

Sums, rearrangements, and norms

Stephen Semmes
Rice University

Abstract

These informal notes deal with a number of questions related to sums and integrals in analysis.

Contents

I	Basic notions	8
1	Real and complex numbers	8
2	Rearrangements	9
3	Generalized convergence	9
4	Nets	10
5	Norms on vector spaces	11
6	Bounded functions	12
7	Summable functions	12
8	p-Summable functions	12
9	Monotonicity	13
10	p-Norms, $0 < p \leq 1$	14
11	Metric spaces	15
12	Infinite series	16
13	$c_0(E)$	17
14	Generalized convergence, 2	17

15 Summable functions, 2	18
16 A special case	18
17 Inner product spaces	19
18 Inner product spaces, 2	20
19 Infinite series, 2	21
20 Hölder's inequality	22
21 Bounded linear functionals	23
22 Hölder's inequality, 2	24
23 Hilbert spaces	24
24 The Hahn–Banach theorem	26
25 Weak summability	26
26 Bounded partial sums	27
27 Bounded finite subsums	28
28 Uniform boundedness	29
29 Uniform boundedness, 2	30
30 Sums and linear functionals	31
31 Seminorms	32
32 Sums in dual spaces	32
33 Seminorms, 2	33
34 Isometric embeddings	34
 II Functions, measures, and paths	 35
35 Uniform boundedness, 3	35
36 Real and complex measures	38
37 Vector-valued measures	41

38 The Radon–Nikodym theorem	44
39 The Lebesgue decomposition	47
40 The Riesz representation theorem	48
41 Lengths of paths	51
42 Lipschitz mappings	52
43 Bounded variation	53
44 Functions and measures	54
45 Continuity conditions	56
46 Maximal functions	57
47 Lebesgue’s theorem	58
48 Singular measures	60
49 Differentiability almost everywhere	61
50 Maximal functions, 2	61
51 Vector-valued functions	63
52 Uniform boundedness, 4	63
53 Weak* derivatives	64
54 Lipschitz functions	65
55 Averages	66
56 L^p derivatives	67
57 L^p Lipschitz conditions	69
58 Dyadic intervals	70
59 Dyadic averages	71
60 Rademacher functions	72
61 L^p estimates	73
62 Rademacher sums	75

63 Lacunary series	76
64 Walsh functions	77
65 Independent random variables	78
66 Linear functions on \mathbb{R}^n	81
67 Countability conditions	82
68 Separation conditions	83
69 Metrizable	84
70 Partitions of unity	85
71 Product spaces	86
72 Product spaces, 2	88
 III Conditional expectation and martingales	 89
73 σ -Subalgebras	89
74 L^p Spaces	91
75 Conditional expectation	91
76 Product spaces, 3	92
77 Measurable partitions	93
78 Basic properties	94
79 Distances between measurable sets	98
80 Sequences of σ -subalgebras	101
81 Martingales	102
82 L^p Boundedness	104
83 Uniform integrability	105
84 Maximal functions, 3	107
85 Convergence almost everywhere	109

86 Other measures	111
87 Finitely-additive measures	113
88 Maximal functions, 4	114
89 Decreasing sequences of σ -algebras	116
90 Doubly-infinite sequences	119
91 Submartingales	120
92 Another variant	123
93 Averaging functions	124
94 Shift mappings	125
95 Families of σ -subalgebras	125
96 Stopping times	126
97 Ultrametrics	130
 IV Vector-valued functions	 131
98 Some randomized sums	132
99 Randomized sums, 2	133
100 The unit square	134
101 Partitions and products	136
102 Partitions and vectors	137
103 Vector-valued martingales	138
104 L^1 -Valued martingales	139
105 Pointwise convergence	140
106 Another scenario	141
107 Hilbert space martingales	143
108 Nonnegative submartingales	144

109 L^p -Valued martingales	145
110 Another criterion	146
111 ℓ^1 -Valued martingales	147
112 Differentiability of paths	149
113 Paths in ℓ^1	150
114 L^p -Valued functions	151
115 Continuous L^p -valued functions	152
116 Lipschitz L^p -valued functions	152
117 More duality	155
118 ℓ^p -Valued functions	156
119 Products and σ -subalgebras	156
120 σ -Subalgebras and vectors	157
121 Martingales and products	158
122 ℓ^p -Valued martingales	160
123 Approximation in product spaces	162
124 Mixed norms	162
125 Mixed-norm martingales	164
126 Mixed-norm convergence	165
127 The ℓ^p version	166
128 The doubling condition	170
129 Paths and martingales	171
130 L^∞ Norms	173
131 Paths and measures	173
132 Paths and integrals	175
133 Integrating vector measures	177

134 Measures and orthogonality	179
135 Paths and orthogonality	180
136 Minkowski's integral inequality	182
137 Spaces of measures	184
138 Products and measures	186
139 L^p -Valued measures	187
140 ℓ^1 -Valued measures	188
141 Finite sums	189
142 Approximations	191
143 Uniform convexity	192
144 Uniform convexity and measures	194
145 Uniform convexity and paths	196
146 Uniform convexity and martingales	198
147 Strict convexity	201
148 Minimizing distances	204
149 Another approximation argument	205
150 Examples in ℓ^p	209
151 Uniform convergence	209
152 Bounded sums	211
153 Bounded coefficients	214
154 Another norm	215
155 Additional properties	217
156 Tori	218
157 Norms and linear functionals	219
158 Sums and $c_0(E)$	220

159 Integrability	221
160 Bounded measures	222
161 Weak* measurability	224
162 Weak* measures	225
163 Weak* integrability	227
References	228

Part I

Basic notions

1 Real and complex numbers

Of course, the real numbers \mathbf{R} are contained in the complex numbers \mathbf{C} , and every $z \in \mathbf{C}$ can be expressed as $z = x + yi$, where $x, y \in \mathbf{R}$ and $i^2 = -1$. In this case, x and y are called the real and imaginary parts of z , respectively. The *complex conjugate* \bar{z} of z is given by

$$(1.1) \quad \bar{z} = x - yi.$$

It is easy to see that

$$(1.2) \quad \overline{z + w} = \bar{z} + \bar{w}$$

and

$$(1.3) \quad \overline{\bar{z} \bar{w}} = z w$$

for every $z, w \in \mathbf{C}$. The *modulus* $|z|$ of z is given by

$$(1.4) \quad |z| = (x^2 + y^2)^{1/2}.$$

Thus

$$(1.5) \quad |z|^2 = z \bar{z}.$$

This implies that

$$(1.6) \quad |z w|^2 = (z w) \overline{z w} = z w \bar{z} \bar{w} = |z|^2 |w|^2$$

for every $z, w \in \mathbf{C}$, and hence

$$(1.7) \quad |z w| = |z| |w|.$$

Note that the modulus of a real number is the same as its absolute value, and that the modulus of $z = x + yi \in \mathbf{C}$ is the same as the Euclidean norm of $(x, y) \in \mathbf{R}^2$.

2 Rearrangements

Let $\sum_{j=1}^{\infty} a_j$ be an infinite series of real or complex numbers. If π is a one-to-one mapping from the set \mathbf{Z}_+ of positive integers onto itself, then the series

$$(2.1) \quad \sum_{j=1}^{\infty} a_{\pi(j)}$$

is said to be a *rearrangement* of $\sum_{j=1}^{\infty} a_j$.

Remember that $\sum_{j=1}^{\infty} a_j$ converges if the sequence of partial sums $\sum_{j=1}^n a_j$ converges as $n \rightarrow \infty$. If a_j is a nonnegative real number for each j , then the partial sums are monotone increasing, and convergence is equivalent to boundedness of the partial sums. In this case, convergence of $\sum_{j=1}^{\infty} a_j$ implies the convergence of every rearrangement (2.1), and the values of these sums are the same. More precisely,

$$(2.2) \quad \sum_{j=1}^n a_{\pi(j)} \leq \sum_{j=1}^N a_j$$

when $\pi(1), \dots, \pi(n) \leq N$, so that the boundedness of the partial sums of $\sum_{j=1}^{\infty} a_j$ implies the boundedness of the partial sums of (2.1). Similarly,

$$(2.3) \quad \sum_{j=1}^n a_j \leq \sum_{j=1}^N a_{\pi(j)}$$

when $\pi^{-1}(1), \dots, \pi^{-1}(n) \leq N$, and these two simple estimates imply that the suprema of the partial sums of $\sum_{j=1}^{\infty} a_j$ and (2.1) are the same.

An infinite series $\sum_{j=1}^{\infty} a_j$ of real or complex numbers is said to converge absolutely if $\sum_{j=1}^{\infty} |a_j|$ converges. It is well known that absolute convergence implies convergence, by the Cauchy criterion. If $\sum_{j=1}^{\infty} a_j$ converges absolutely, then the preceding discussion implies that (2.1) also converges absolutely, and one can show that the two sums have the same value. This is trivial when $a_j = 0$ for all but finitely many j , and otherwise $\sum_{j=1}^{\infty} a_j$ can be approximated by series with this property. Alternatively, $\sum_{j=1}^{\infty} a_j$ may be expressed as a linear combination of convergent series whose terms are nonnegative real numbers, so that the equality of the sums reduces to the previous case.

3 Generalized convergence

Let E be a nonempty set, and let $f(x)$ be a real or complex-valued function on E . Let us say that $\sum_{x \in E} f(x)$ converges in the generalized sense if there is a $\lambda \in \mathbf{R}$ or \mathbf{C} , as appropriate, such that for each $\epsilon > 0$ there is a finite set $A_{\epsilon} \subseteq E$ for which

$$(3.1) \quad \left| \sum_{x \in B} f(x) - \lambda \right| < \epsilon$$

whenever $B \subseteq E$ is a finite set that satisfies $A_\epsilon \subseteq B$. It is easy to see that such a λ is unique when it exists, in which case $\sum_{x \in E} f(x)$ is defined to be λ .

If $f(x)$ has this property and π is a one-to-one mapping of E onto itself, then $f(\pi(x))$ has the same property, and

$$(3.2) \quad \sum_{x \in E} f(\pi(x)) = \sum_{x \in E} f(x).$$

This follows from the fact that

$$(3.3) \quad \sum_{x \in A} f(\pi(x)) = \sum_{x \in \pi(A)} f(x)$$

for every finite set $A \subseteq E$. Thus this definition of $\sum_{x \in E} f(x)$ is automatically invariant under rearrangements.

Suppose that $f(x)$ is a nonnegative real number for each $x \in E$. If the partial sums $\sum_{x \in A} f(x)$ over finite subsets A of E are uniformly bounded, then $\sum_{x \in E} f(x)$ converges in the generalized sense, and

$$(3.4) \quad \sum_{x \in E} f(x) = \sup \left\{ \sum_{x \in A} f(x) : A \subseteq E \text{ has only finitely many elements} \right\}.$$

If $f(x)$ is a real or complex-valued function on E such that the sums $\sum_{x \in A} |f(x)|$ over finite sets $A \subseteq E$ are bounded, then $\sum_{x \in E} f(x)$ also converges in the generalized sense. This follows by expressing $f(x)$ as a linear combination of nonnegative real-valued functions for which the partial sums over finite subsets of E are bounded.

Conversely, if $f(x)$ is a real or complex-valued function on E such that $\sum_{x \in E} f(x)$ converges in the generalized sense, then the sums $\sum_{x \in A} |f(x)|$ over finite subsets A of E are uniformly bounded. To see this, one can take $\epsilon = 1$ in the definition of convergence to get a finite set $A_1 \subseteq E$ for which the partial sums $\sum_{x \in B} f(x)$ over finite subsets B of E with $A_1 \subseteq B$ are uniformly bounded. This implies that the partial sums $\sum_{x \in A} f(x)$ over arbitrary finite sets $A \subseteq E$ are bounded, by taking $B = A \cup A_1$, and using the fact that the sums over subsets of A_1 are bounded. The boundedness of the partial sums of $|f(x)|$ can then be obtained by applying this to finite sets $A \subseteq E$ on which $f(x)$ is positive or negative in the real case, or on which the real or imaginary parts of $f(x)$ are positive or negative in the complex case.

4 Nets

A partially ordered set (A, \prec) is said to be a *directed system* if for every $a, b \in A$ there is a $c \in A$ such that $a, b \prec c$. A *net* $\{x_a\}_{a \in A}$ indexed by A assigns to each $a \in A$ an element x_a of a set X . If X is a topological space, then the net $\{x_a\}_{a \in A}$ converges to $x \in X$ if for every open set $U \subseteq X$ with $x \in U$ there is an $a \in A$ such that $x_b \in U$ when $b \in A$ and $a \prec b$. This reduces to the usual

definition of convergence of a sequence when A is the set of positive integers with the standard ordering. Now let E be a nonempty set, and let $f(x)$ be a real or complex-valued function on E . The collection of nonempty finite subsets of E is partially ordered by inclusion, and defines a directed system. More precisely, any two finite subsets of E is contained in their union, which is also a finite subset of E . Consider the net associated to this directed system that assigns to each nonempty finite set $B \subseteq E$ the real or complex number $\sum_{x \in B} f(x)$. It is easy to see that convergence of this net in \mathbf{R} or \mathbf{C} , as appropriate, is the same as convergence of $\sum_{x \in E} f(x)$ in the sense described in the previous section.

5 Norms on vector spaces

Let V be a vector space over the real or complex numbers. A *norm* on V is a nonnegative real-valued function $\|v\|$ defined for $v \in V$ such that $\|v\| = 0$ if and only if $v = 0$,

$$(5.1) \quad \|tv\| = |t| \|v\|$$

for every $v \in V$ and $t \in \mathbf{R}$ or \mathbf{C} , as appropriate, and

$$(5.2) \quad \|v + w\| \leq \|v\| + \|w\|$$

for every $v, w \in V$.

A set $E \subseteq V$ is said to be *convex* if for every $v, w \in E$ and $t \in \mathbf{R}$ with $0 < t < 1$,

$$(5.3) \quad tv + (1 - t)w \in E.$$

If $\|v\|$ is a norm on V and

$$(5.4) \quad B_1 = \{v \in V : \|v\| \leq 1\}$$

is the corresponding closed unit ball, then it is easy to see that B_1 is a convex set in V .

Conversely, suppose that $\|v\|$ is a nonnegative real-valued function on V that satisfies the positivity condition $\|v\| > 0$ when $v \neq 0$ and the homogeneity condition (5.1). If B_1 is convex, then one can show that $\|v\|$ satisfies the triangle inequality (5.2), and hence that $\|v\|$ is a norm. To see this, let $v, w \in V$ be given, with $v, w \neq 0$, since otherwise (5.2) is trivial. Put

$$(5.5) \quad v' = \frac{v}{\|v\|}, \quad w' = \frac{w}{\|w\|},$$

so that $\|v'\| = \|w'\| = 1$. Thus $v', w' \in B_1$, and hence

$$(5.6) \quad \|tv' + (1 - t)w'\| \leq 1$$

when $t \in \mathbf{R}$ and $0 < t < 1$, by hypothesis. If $t = \|v\|/(\|v\| + \|w\|)$, then $1 - t = \|w\|/(\|v\| + \|w\|)$, and

$$(5.7) \quad tv' + (1 - t)w' = \frac{v + w}{\|v\| + \|w\|}.$$

Therefore

$$(5.8) \quad \left\| \frac{v+w}{\|v\| + \|w\|} \right\| \leq 1,$$

which implies (5.2), as desired.

6 Bounded functions

Let E be a nonempty set, and consider the spaces $\ell^\infty(E, \mathbf{R})$, $\ell^\infty(E, \mathbf{C})$ of real or complex-valued functions on E that are bounded. It is sometimes convenient to use the notation $\ell^\infty(E)$ to refer to either of these spaces, which are vector spaces with respect to pointwise addition and scalar multiplication. The supremum or ℓ^∞ norm is defined as usual by

$$(6.1) \quad \|f\|_\infty = \sup\{|f(x)| : x \in E\}.$$

It is easy to see that this is a norm on $\ell^\infty(E)$, because of the triangle inequality for the ordinary absolute value on \mathbf{R} or modulus on \mathbf{C} .

7 Summable functions

A real or complex-valued function $f(x)$ on a nonempty set E is said to be *summable* if the partial sums $\sum_{x \in A} |f(x)|$ over nonempty finite subsets A of E are uniformly bounded. This is equivalent to the convergence of $\sum_{x \in E} |f(x)|$ in the sense of Section 3, whose value is equal to the supremum of $\sum_{x \in A} |f(x)|$ over all nonempty finite sets $A \subseteq E$. Let $\ell^1(E, \mathbf{R})$, $\ell^1(E, \mathbf{C})$ be the spaces of summable real or complex-valued functions on E , respectively, which may also be denoted by $\ell^1(E)$ to include both cases at the same time. It is easy to see that these are vector spaces with respect to pointwise addition and scalar multiplication, and that

$$(7.1) \quad \|f\|_1 = \sum_{x \in E} |f(x)|$$

defines a norm on these spaces.

8 p -Summable functions

Let $f(x)$ be a real or complex-valued function on a nonempty set E , and let p be a positive real number. If $|f(x)|^p$ is a summable function on E , then we say that $f(x)$ is *p -summable* on E . The spaces of real or complex-valued p -summable functions on E are denoted $\ell^p(E, \mathbf{R})$, $\ell^p(E, \mathbf{C})$, respectively, or simply $\ell^p(E)$ to include both cases at the same time. One can check that these are vector spaces over the real or complex numbers, as appropriate, with respect to pointwise addition and scalar multiplication of functions.

If f is a p -summable function on E , then put

$$(8.1) \quad \|f\|_p = \left(\sum_{x \in E} |f(x)|^p \right)^{1/p}.$$

This satisfies the positivity and homogeneity properties of a norm on $\ell^p(E)$ for every $p > 0$. Let us check that this is a norm on $\ell^p(E)$ when $p \geq 1$. As in Section 5, it suffices to show that the closed unit ball in $\ell^p(E)$ associated to $\|f\|_p$ is convex when $p \geq 1$. Equivalently, if f, g are p -summable functions on E such that $\|f\|_p, \|g\|_p \leq 1$, then we would like to check that

$$(8.2) \quad \|t f + (1 - t) g\|_p \leq 1$$

when $t \in \mathbf{R}$ and $0 < t < 1$. The main point is that

$$(8.3) \quad \begin{aligned} |t f(x) + (1 - t) g(x)|^p &\leq (t |f(x)| + (1 - t) |g(x)|)^p \\ &\leq t |f(x)|^p + (1 - t) |g(x)|^p \end{aligned}$$

for every $x \in E$, because of the convexity of the function $\phi_p(r) = r^p$ on the nonnegative real numbers when $p \geq 1$. Hence

$$(8.4) \quad \sum_{x \in E} |t f(x) + (1 - t) g(x)|^p \leq t \sum_{x \in E} |f(x)|^p + (1 - t) \sum_{x \in E} |g(x)|^p \leq 1.$$

9 Monotonicity

Let p be a positive real number, and let f be a real or complex-valued p -summable function on a nonempty set E . Clearly

$$(9.1) \quad |f(x)| \leq \|f\|_p$$

for every $x \in E$, which implies that f is bounded and satisfies

$$(9.2) \quad \|f\|_\infty \leq \|f\|_p.$$

If $q \geq p$, then f is also q -summable, because

$$(9.3) \quad |f(x)|^q \leq \|f\|_\infty^{q-p} |f(x)|^p \leq \|f\|_p^{q-p} |f(x)|^p$$

for every $x \in E$. Moreover,

$$(9.4) \quad \|f\|_q^q = \sum_{x \in E} |f(x)|^q \leq \|f\|_p^{q-p} \sum_{x \in E} |f(x)|^p = \|f\|_p^q,$$

and hence

$$(9.5) \quad \|f\|_q \leq \|f\|_p.$$

If $q = 1$, then we get that

$$(9.6) \quad \left(\sum_{x \in E} |f(x)| \right)^p \leq \sum_{x \in E} |f(x)|^p$$

when f is p -summable and $0 < p \leq 1$. In particular,

$$(9.7) \quad (a + b)^p \leq a^p + b^p$$

for every pair of nonnegative real numbers a, b when $0 < p \leq 1$, by applying the previous inequality to a set E with exactly two elements. Conversely, one can apply (9.7) repeatedly to get

$$(9.8) \quad \left(\sum_{j=1}^n a_j \right)^p \leq \sum_{j=1}^n a_j^p$$

for any positive integer n and nonnegative real numbers a_1, \dots, a_n , which implies the analogous inequality for arbitrary sums by passing to a suitable limit.

10 p -Norms, $0 < p \leq 1$

Let V be a vector space over the real or complex numbers, and let $\|v\|$ be a nonnegative real-valued function on V such that $\|v\| > 0$ when $v \neq 0$ and

$$(10.1) \quad \|tv\| = |t| \|v\|$$

for every $v \in V$ and $t \in \mathbf{R}$ or \mathbf{C} , as appropriate. We say that $\|v\|$ is a p -norm, $0 < p \leq 1$, if in addition

$$(10.2) \quad \|v + w\|^p \leq \|v\|^p + \|w\|^p$$

for every $v, w \in V$. This reduces to the ordinary triangle inequality (5.2) when $p = 1$, so that a 1-norm is the same as a norm. For example, $\|f\|_p$ defines a p -norm on $\ell^p(E)$ for any nonempty set E when $0 < p \leq 1$, because of (9.7).

Equivalently, $\|v\|$ is a p -norm when

$$(10.3) \quad \|v + w\| \leq (\|v\|^p + \|w\|^p)^{1/p}$$

for every $v, w \in V$. As in the previous section, the right side of this inequality is monotone decreasing in p . Hence a p -norm is also a \tilde{p} -norm when $0 < \tilde{p} \leq p \leq 1$.

Let B_1 be the closed unit ball associated to $\|v\|$, as in (5.4). If $\|v\|$ is a p -norm, then

$$(10.4) \quad av + bw \in B_1$$

whenever $v, w \in B_1$ and a, b are nonnegative real numbers such that $a^p + b^p \leq 1$. Conversely, let us check that this property implies that $\|v\|$ is a p -norm, as in Section 5. Let v, w be nonzero vectors in V , and put $v' = v/\|v\|$, $w' = w/\|w\|$, as before. Also put

$$(10.5) \quad a = \frac{\|v\|}{(\|v\|^p + \|w\|^p)^{1/p}}, \quad b = \frac{\|w\|}{(\|v\|^p + \|w\|^p)^{1/p}}.$$

Thus

$$(10.6) \quad a^p + b^p = \frac{\|v\|^p}{\|v\|^p + \|w\|^p} + \frac{\|w\|^p}{\|v\|^p + \|w\|^p} = 1,$$

and hence

$$(10.7) \quad a v' + b w' = \frac{v + w}{(\|v\|^p + \|w\|^p)^{1/p}} \in B_1.$$

This implies the p -norm version of the triangle inequality when $v, w \neq 0$, and of course it is trivial when v or w is equal to 0.

11 Metric spaces

Remember that a metric space is a set M with a nonnegative real-valued function $d(x, y)$ defined for $x, y \in M$ such that $d(x, y) = 0$ if and only if $x = y$,

$$(11.1) \quad d(y, x) = d(x, y)$$

for every $x, y \in M$, and

$$(11.2) \quad d(x, z) \leq d(x, y) + d(y, z)$$

for every $x, y, z \in M$. If V is a real or complex vector space equipped with a norm $\|v\|$, then

$$(11.3) \quad d(v, w) = \|v - w\|$$

is a metric on V . Similarly, if $\|v\|$ is a p -norm on V for some p , $0 < p \leq 1$, then

$$(11.4) \quad d(v, w) = \|v - w\|^p$$

is a metric on V .

Let $(M, d(x, y))$ be a metric space. A sequence $\{x_j\}_{j=1}^{\infty}$ of elements of M is said to *converge* to $x \in M$ if for every $\epsilon > 0$ there is an $L \geq 1$ such that

$$(11.5) \quad d(x_j, x) < \epsilon$$

for every $j \geq L$. We say that $\{x_j\}_{j=1}^{\infty}$ is a *Cauchy sequence* if for every $\epsilon > 0$ there is an $L \geq 1$ such that

$$(11.6) \quad d(x_j, x_l) < \epsilon$$

for every $j, l \geq L$. It is easy to check that every convergent sequence is a Cauchy sequence, and a metric space is said to be *complete* if every Cauchy sequence converges to an element of the space. For example, it is well known that the real and complex numbers are complete with respect to their standard metrics.

If $\{x_j\}_{j=1}^{\infty}$ is a sequence of elements of M with the property that

$$(11.7) \quad \sum_{j=1}^{\infty} d(x_j, x_{j+1})$$

converges, then $\{x_j\}_{j=1}^{\infty}$ is a Cauchy sequence in M . This uses the triangle inequality to get that

$$(11.8) \quad d(x_k, x_l) \leq \sum_{j=k}^{l-1} d(x_j, x_{j+1})$$

when $k < l$. If M is complete, then it follows that $\{x_j\}_{j=1}^\infty$ converges in M . Conversely, if $\{x_j\}_{j=1}^\infty$ is a Cauchy sequence in M , then there is a subsequence $\{x_{j_n}\}_{n=1}^\infty$ of $\{x_j\}_{j=1}^\infty$ such that

$$(11.9) \quad d(x_{j_n}, x_{j_{n+1}}) \leq 2^{-n}$$

for each n , which implies that

$$(11.10) \quad \sum_{n=1}^{\infty} d(x_{j_n}, x_{j_{n+1}})$$

converges. If this subsequence converges, then $\{x_j\}_{j=1}^\infty$ converges to the same limit, because it is a Cauchy sequence.

Let E be a nonempty set, and consider $\ell^p(E)$, $0 < p \leq \infty$. This is a metric space with respect to the metric associated to the norm $\|f\|_p$ when $p \geq 1$, or the p -norm $\|f\|_p$ when $0 < p \leq 1$, and it is well known that this space is complete. For if $\{f_j\}_{j=1}^\infty$ is a Cauchy sequence in $\ell^p(E)$, then it is easy to see that $\{f_j(x)\}_{j=1}^\infty$ is a Cauchy sequence in \mathbf{R} or \mathbf{C} for each $x \in E$, as appropriate. This implies that $\{f_j(x)\}_{j=1}^\infty$ converges pointwise on E , since the real and complex numbers are complete. One can check that the limit $f(x)$ is also in $\ell^p(E)$, and that $\{f_j\}_{j=1}^\infty$ converges to f in the ℓ^p metric, as desired.

12 Infinite series

Let V be a real or complex vector space equipped with a norm or p -norm $\|v\|$, $0 < p \leq 1$. This determines a natural metric on V , as in the previous section. As usual, an infinite series $\sum_{j=1}^\infty v_j$ with terms $v_j \in V$ is said to converge if the corresponding sequence of partial sums $\sum_{j=1}^n v_j$ converges in V as $n \rightarrow \infty$. Let us say that $\sum_{j=1}^\infty v_j$ converges absolutely if

$$(12.1) \quad \sum_{j=1}^{\infty} \|v_j\|$$

converges when $\|v\|$ is a norm, and if

$$(12.2) \quad \sum_{j=1}^{\infty} \|v_j\|^p$$

converges when $\|v\|$ is a p -norm. Note that the convergence of (12.2) is more restrictive as p decreases, as in Section 9. As in the previous section, absolute convergence of $\sum_{j=1}^\infty v_j$ implies that the sequence of partial sums $\sum_{j=1}^n v_j$ is a Cauchy sequence. In particular, absolute convergence implies convergence when V is complete. Conversely, V is complete if every absolutely convergent series with terms in V converges in V , by another argument mentioned in the previous section.

13 $c_0(E)$

Let E be a nonempty set, and let $f(x)$ be a real or complex-valued function on E . We say that f vanishes at infinity on E if for every $\epsilon > 0$, $|f(x)| \geq \epsilon$ for only finitely many $x \in E$. The spaces of real or complex-valued functions on E that vanish at infinity are denoted $c_0(E, \mathbf{R})$, $c_0(E, \mathbf{C})$, respectively, and are vector spaces with respect to pointwise addition and scalar multiplication of functions. As usual, we may also use $c_0(E)$ to refer to both cases at the same time. Note that $f(x) \neq 0$ for only finitely or countably many $x \in E$ when $f \in c_0(E)$.

If f vanishes at infinity on E , then f is bounded, and so $c_0(E)$ is a linear subspace of $\ell^\infty(E)$. More precisely, one can check that $c_0(E)$ is a closed linear subspace of $\ell^\infty(E)$ with respect to the ℓ^∞ norm. A function f on E is said to have finite support if $f(x) \neq 0$ for only finitely many $x \in E$, in which case it obviously vanishes at infinity. One can also check that functions with finite support are dense in $c_0(E)$ with respect to the ℓ^∞ norm, so that $c_0(E)$ is the same as the closure in $\ell^\infty(E)$ of the linear subspace of functions with finite support.

If a function f on E is p -summable for some $p > 0$, then f vanishes at infinity on E . More precisely, the number of $x \in E$ such that $|f(x)| \geq \epsilon$ is less than or equal to

$$(13.1) \quad \epsilon^{-p} \sum_{x \in E} |f(x)|^p.$$

Of course, a function f with finite support on E is p -summable for every $p > 0$. It is not difficult to show that functions with finite support on E are dense in $\ell^p(E)$ when $0 < p < \infty$.

14 Generalized convergence, 2

Let E be a nonempty set, let V be a real or complex vector space with a norm or p -norm $\|v\|$, $0 < p \leq 1$, and let $f(x)$ be a V -valued function on E . We say that $\sum_{x \in E} f(x)$ converges in the generalized sense if there is a $\lambda \in V$ such that for every $\epsilon > 0$ there is a finite set $A_\epsilon \subseteq E$ such that

$$(14.1) \quad \left\| \sum_{x \in B} f(x) - \lambda \right\| < \epsilon$$

whenever $B \subseteq E$ is a finite set that satisfies $A_\epsilon \subseteq B$. It is easy to see that λ is unique when it exists, in which case it may be denoted $\sum_{x \in E} f(x)$. Of course, this is the same as the definition in Section 3 when $V = \mathbf{R}$ or \mathbf{C} , and it is equivalent to the convergence of the net of partial sums of $f(x)$ over finite subsets of E as in Section 4.

Similarly, we say that $\sum_{x \in E} f(x)$ satisfies the generalized Cauchy criterion if for every $\epsilon > 0$ there is a finite set $A_\epsilon \subseteq E$ such that

$$(14.2) \quad \left\| \sum_{x \in B} f(x) \right\| < \epsilon$$

whenever $B \subseteq E$ is a finite set with $A_\epsilon \cap B = \emptyset$. If $\sum_{x \in E} f(x)$ converges in the generalized sense, then it is easy to see that $\sum_{x \in E} f(x)$ satisfies the generalized Cauchy criterion. Conversely, let us check that $\sum_{x \in E} f(x)$ converges in the generalized sense when $\sum_{x \in E} f(x)$ satisfies the generalized Cauchy criterion and V is complete.

If $\sum_{x \in E} f(x)$ satisfies the generalized Cauchy criterion, then it is easy to see that $\|f(x)\|$ vanishes at infinity on E , by considering sets $B \subseteq E$ with only one element in the previous definition. In particular, $f(x) \neq 0$ for only finitely or countably many $x \in E$. If $f(x) \neq 0$ for only finitely many $x \in E$, then convergence of the sum is trivial, and so we suppose that $f(x) \neq 0$ for countably many x . Let $\{x_j\}_{j=1}^\infty$ be an enumeration of the set of $x \in E$ such that $f(x) \neq 0$, so that each element of this set occurs in the sequence exactly once, and consider the infinite series $\sum_{j=1}^\infty f(x_j)$. Using the generalized Cauchy criterion for $\sum_{x \in E} f(x)$, one can check that the sequence of partial sums of $\sum_{j=1}^\infty f(x_j)$ forms a Cauchy sequence in V . If V is complete, then it follows that $\sum_{j=1}^\infty f(x_j)$ converges in V . Using the generalized Cauchy criterion for $\sum_{x \in E} f(x)$ again, one can show that $\sum_{x \in E} f(x)$ converges in the generalized sense, and that the sum is the same as $\sum_{j=1}^\infty f(x_j)$.

15 Summable functions, 2

Let E be a nonempty set, and let V be a real or complex vector space equipped with a norm or p -norm $\|v\|$ for $0 < p \leq 1$. Suppose that f is a V -valued function on E such that $\|f(x)\|$ is summable on E when $\|v\|$ is a norm on V , or that $\|f(x)\|^p$ is summable on E when $\|v\|$ is a p -norm, $0 < p \leq 1$. If $B \subseteq E$ is a finite set, then we have that

$$(15.1) \quad \left\| \sum_{x \in B} f(x) \right\| \leq \sum_{x \in B} \|f(x)\|$$

in the first case, and

$$(15.2) \quad \left\| \sum_{x \in B} f(x) \right\|^p \leq \sum_{x \in B} \|f(x)\|^p$$

in the second case. In both cases, one can use these simple estimates to check that $\sum_{x \in E} f(x)$ satisfies the generalized Cauchy criterion. If V is complete, then it follows that $\sum_{x \in E} f(x)$ converges in the generalized sense, as in the previous section.

16 A special case

Let E be a nonempty set, and suppose that $\phi \in \ell^p(E)$ for some p , $0 < p \leq \infty$. For each $x \in E$, let $\delta_x(y)$ be the function on E defined by $\delta_x(x) = 1$ and $\delta_x(y) = 0$ when $y \neq x$. Consider

$$(16.1) \quad f(x) = \phi(x) \delta_x,$$

as a function on E with values in $\ell^p(E)$. Observe that

$$(16.2) \quad \sum_{x \in E} f(x)(y) = \sum_{x \in E} \phi(x) \delta_x(y) = \phi(y)$$

for each $y \in E$, where these are sums over $x \in E$ of real or complex numbers that are equal to 0 when $x \neq y$ and hence converge trivially. One can also ask about the convergence of $\sum_{x \in E} f(x)$ in the generalized sense to ϕ , as a sum of elements of $\ell^p(E)$. Of course,

$$(16.3) \quad \|f(x)\|_p = |\phi(x)| \|\delta_x\|_p = |\phi(x)|$$

for every $x \in E$. Thus $\|f(x)\|_p$ is p -summable on E when $0 < p < \infty$, and bounded on E when $p = \infty$. If $0 < p \leq 1$, then this is the same as the summability condition mentioned in the previous section. However, one can check that $\sum_{x \in E} f(x)$ converges to ϕ in the generalized sense in $\ell^p(E)$ for every positive real number p . If $p = \infty$, then $\sum_{x \in E} f(x)$ converges to ϕ in the generalized sense in $\ell^\infty(E)$ if and only if $\phi \in c_0(E)$.

17 Inner product spaces

An *inner product* on a real or complex vector space V is a real or complex-valued function $\langle v, w \rangle$, as appropriate, defined for $v, w \in V$ and satisfying the following three conditions. First, $\langle v, w \rangle$ is a linear function of v for each $w \in W$. Second,

$$(17.1) \quad \langle w, v \rangle = \langle v, w \rangle$$

for every $v, w \in V$ in the real case, and

$$(17.2) \quad \langle w, v \rangle = \overline{\langle v, w \rangle}$$

in the complex case. In particular,

$$(17.3) \quad \langle v, v \rangle = \overline{\langle v, v \rangle} \in \mathbf{R}$$

for every $v \in V$ in the complex case. Third,

$$(17.4) \quad \langle v, v \rangle > 0$$

for every $v \in V$ with $v \neq 0$.

Put

$$(17.5) \quad \|v\| = \langle v, v \rangle^{1/2}.$$

The *Cauchy-Schwarz inequality* states that

$$(17.6) \quad |\langle v, w \rangle| \leq \|v\| \|w\|$$

for every $v, w \in V$. Using this, one can show that

$$(17.7) \quad \|v + w\| \leq \|v\| + \|w\|$$

for every $v, w \in V$, so that $\|v\|$ defines a norm on V . If V is complete with respect to this norm, then V is said to be a Hilbert space.

Let E be a nonempty set, and let $f, g \in \ell^2(E)$ be given. Remember that

$$(17.8) \quad ab \leq \frac{a^2 + b^2}{2}$$

for every $a, b \geq 0$, since $(a - b)^2 \geq 0$, so that

$$(17.9) \quad \sum_{x \in E} |f(x)| |g(x)| \leq \frac{1}{2} \sum_{x \in E} |f(x)|^2 + \frac{1}{2} \sum_{x \in E} |g(x)|^2 < +\infty.$$

Thus $|f(x)| |g(x)|$ is summable on E , and it is easy to see that

$$(17.10) \quad \langle f, g \rangle = \sum_{x \in E} f(x) g(x)$$

defines an inner product on $\ell^2(E, \mathbf{R})$, and that

$$(17.11) \quad \langle f, g \rangle = \sum_{x \in E} f(x) \overline{g(x)}$$

defines an inner product on $\ell^2(E, \mathbf{C})$. The corresponding norm is the same as the ℓ^2 norm discussed in Section 8. These spaces are also complete, as in Section 11, and are therefore Hilbert spaces.

A pair of vectors v, w in an inner product space V are said to be *orthogonal* if

$$(17.12) \quad \langle v, w \rangle = 0.$$

This may also be expressed by $v \perp w$. In this case,

$$(17.13) \quad \|v + w\|^2 = \langle v + w, v + w \rangle = \langle v, v \rangle + \langle w, w \rangle = \|v\|^2 + \|w\|^2.$$

If $v_1, \dots, v_n \in V$ and $v_j \perp v_l$ when $j \neq l$, then we get that

$$(17.14) \quad \left\| \sum_{j=1}^n v_j \right\|^2 = \sum_{j=1}^n \|v_j\|^2.$$

18 Inner product spaces, 2

Let E be a nonempty set, let $(V, \langle v, w \rangle)$ be an inner product space, and let f be a V -valued function on E such that

$$(18.1) \quad f(x) \perp f(y)$$

when $x \neq y$. Thus

$$(18.2) \quad \left\| \sum_{x \in B} f(x) \right\|^2 = \sum_{x \in B} \|f(x)\|^2$$

for every finite set $B \subseteq E$, as in the previous section. If $\|f(x)\|^2$ is a summable function on E , then it follows that $\sum_{x \in E} f(x)$ satisfies the generalized Cauchy criterion, and hence converges in the generalized sense when V is complete. In this case, one can also check that

$$(18.3) \quad \left\| \sum_{x \in E} f(x) \right\|^2 = \sum_{x \in E} \|f(x)\|^2.$$

19 Infinite series, 2

Let V be a real or complex vector space equipped with a norm or p -norm $\|v\|$, $0 < p \leq 1$, and let $\sum_{j=1}^{\infty} v_j$ be an infinite series with terms in V . This can also be considered as a sum over $E = \mathbf{Z}_+$, so that the notions of convergence in the generalized sense and the generalized Cauchy criterion are applicable. If $\sum_{j=1}^{\infty} v_j$ converges in the ordinary sense and satisfies the generalized Cauchy criterion as a sum over \mathbf{Z}_+ , then it is easy to see that $\sum_{j=1}^{\infty} v_j$ converges in the generalized sense, and to the same sum.

Suppose that $\sum_{j=1}^{\infty} v_j$ does not satisfy the generalized Cauchy criterion. This means that there is an $\epsilon > 0$ such that for each finite set $A \subseteq \mathbf{Z}_+$ there is another finite set $B \subseteq \mathbf{Z}_+$ such that $A \cap B = \emptyset$ and

$$(19.1) \quad \left\| \sum_{j \in B} v_j \right\| \geq \epsilon.$$

Using this repeatedly, one can get finite subsets A_n, B_n of \mathbf{Z}_+ such that $\{1, \dots, n\} \subseteq A_n$, $A_n \cap B_n = \emptyset$, $A_n \cup B_n \subseteq A_{n+1}$, and

$$(19.2) \quad \left\| \sum_{j \in B_n} v_j \right\| \geq \epsilon$$

for each n . Let k_n be the number of elements of A_n and l_n be the number of elements of B_n , so that $n \leq k_n < k_n + l_n \leq k_{n+1}$ for each n . Also let π be a one-to-one mapping of \mathbf{Z}_+ onto itself such that $A_n = \{\pi(1), \dots, \pi(k_n)\}$ and $B_n = \{\pi(k_n + 1), \dots, \pi(k_n + l_n)\}$ for each n . This is easy to arrange, because of the inclusion and disjointness properties of the A_n 's and B_n 's. Thus

$$(19.3) \quad \left\| \sum_{j=k_n+1}^{k_n+l_n} v_j \right\| \geq \epsilon$$

for each n . This implies that the partial sums of $\sum_{j=1}^{\infty} v_{\pi(j)}$ do not form a Cauchy sequence, and in particular that $\sum_{j=1}^{\infty} v_{\pi(j)}$ does not converge in the ordinary sense.

If $\sum_{j=1}^{\infty} v_j$ satisfies the generalized Cauchy criterion, then it is easy to see that the partial sums of every rearrangement $\sum_{j=1}^{\infty} v_{\pi(j)}$ of $\sum_{j=1}^{\infty} v_j$ form a Cauchy sequence. Conversely, if the partial sums of every rearrangement of

$\sum_{j=1}^{\infty} v_j$ form a Cauchy sequence, then $\sum_{j=1}^{\infty} v_j$ satisfies the generalized Cauchy criterion, by the argument in the preceding paragraph. Similarly, every rearrangement of $\sum_{j=1}^{\infty} v_j$ converges to the same sum when $\sum_{j=1}^{\infty} v_j$ converges in the generalized sense. Conversely, if every rearrangement of $\sum_{j=1}^{\infty} v_j$ converges, then $\sum_{j=1}^{\infty} v_j$ satisfies the generalized Cauchy criterion, by the previous remarks. Hence $\sum_{j=1}^{\infty} v_j$ converges in the generalized sense, because it converges in the ordinary sense, as mentioned at the beginning of the section.

20 Hölder's inequality

Let E be a nonempty set, and suppose that $1 \leq p, q \leq \infty$ are *conjugate exponents* in the sense that

$$(20.1) \quad \frac{1}{p} + \frac{1}{q} = 1.$$

If $f \in \ell^p(E)$ and $g \in \ell^q(E)$, then *Hölder's inequality* states that $f g \in \ell^1(E)$, and that

$$(20.2) \quad \|f g\|_1 \leq \|f\|_p \|g\|_q.$$

This is quite straightforward when $p = 1$, $q = \infty$ or $p = \infty$, $q = 1$, and so we focus now on the case where $1 < p, q < \infty$. Note that the $p = q = 2$ case is another version of the Cauchy–Schwarz inequality.

If a, b are nonnegative real numbers, then

$$(20.3) \quad a b \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

This can be seen as a consequence of the convexity of the exponential function. In particular,

$$(20.4) \quad |f(x)| |g(x)| \leq \frac{|f(x)|^p}{p} + \frac{|g(x)|^q}{q}$$

for every $x \in E$. Hence

$$(20.5) \quad \sum_{x \in B} |f(x)| |g(x)| \leq \frac{1}{p} \sum_{x \in B} |f(x)|^p + \frac{1}{q} \sum_{x \in B} |g(x)|^q \leq \frac{\|f\|_p^p}{p} + \frac{\|g\|_q^q}{q}$$

for every finite set $B \subseteq E$. This implies that $f g$ is summable on E , with

$$(20.6) \quad \|f g\|_1 \leq \frac{\|f\|_p^p}{p} + \frac{\|g\|_q^q}{q}.$$

This implies Hölder's inequality when $\|f\|_p = \|g\|_q = 1$. Otherwise, if $f, g \neq 0$, then we can apply this to

$$(20.7) \quad \tilde{f} = \frac{f}{\|f\|_p}, \quad \tilde{g} = \frac{g}{\|g\|_q}.$$

Thus $\tilde{f} \in \ell^p(E)$, $\tilde{g} \in \ell^q(E)$, $\|\tilde{f}\|_p = \|\tilde{g}\|_q = 1$, and the previous inequality implies that

$$(20.8) \quad \frac{\|f g\|_1}{\|f\|_p \|g\|_q} = \|\tilde{f} \tilde{g}\|_1 \leq 1.$$

Of course, Hölder's inequality is trivial when either f or g is identically 0 on E .

21 Bounded linear functionals

Let V be a vector space over the real or complex numbers. A *linear functional* on V is simply a linear mapping from V into \mathbf{R} or \mathbf{C} , as appropriate. Suppose now that V is also equipped with a norm $\|v\|$. A linear functional λ on V is said to be *bounded* with respect to this norm if there is a nonnegative real number C such that

$$(21.1) \quad |\lambda(v)| \leq C \|v\|$$

for every $v \in V$. In this case, we put

$$(21.2) \quad \|\lambda\|_* = \sup\{|\lambda(v)| : v \in V, \|v\| \leq 1\},$$

which is the same as the smallest $C \geq 0$ for which the previous inequality holds.

The boundedness of a linear functional λ on V implies that

$$(21.3) \quad |\lambda(v) - \lambda(w)| = |\lambda(v - w)| \leq C \|v - w\|$$

for some $C \geq 0$ and every $v, w \in V$. This shows that a bounded linear functional λ is uniformly continuous on V . Conversely, if a linear functional λ on V is continuous at 0, then there is a $\delta > 0$ such that

$$(21.4) \quad |\lambda(v)| < 1$$

for every $v \in V$ with $\|v\| < \delta$. This implies that λ is bounded, with $C = 1/\delta$.

The space of arbitrary linear functionals on V is a vector space with respect to pointwise addition and scalar multiplication of functions. It is easy to see that the space V^* of bounded linear functionals on V is also a vector space in this way, and that $\|\lambda\|_*$ defines a norm on V^* , known as the *dual norm*. Note that V^* is automatically complete with respect to the dual norm. For if $\{\lambda_j\}_{j=1}^\infty$ is a Cauchy sequence of bounded linear functionals on V with respect to the dual norm, then $\{\lambda_j(v)\}_{j=1}^\infty$ is a Cauchy sequence of real or complex numbers, as appropriate, for each $v \in V$. Hence $\{\lambda_j(v)\}_{j=1}^\infty$ converges in \mathbf{R} or \mathbf{C} for each $v \in V$, by completeness. It is easy to see that the limit defines a linear functional λ on V , which is also bounded because the λ_j 's have uniformly bounded dual norms. One can also show that $\{\lambda_j\}_{j=1}^\infty$ converges to λ with respect to the dual norm, using the fact that $\{\lambda_j\}_{j=1}^\infty$ is a Cauchy sequence with respect to the dual norm.

The definitions of bounded linear functionals and the dual norm also make sense when $\|v\|$ is a p -norm on V . The dual space V^* is still a vector space in this case, and the dual norm is still a norm on V^* , and not just a p -norm. The dual space is also complete with respect to the dual norm, but there are some other problems with the dual space when $\|v\|$ is not a norm, as we shall see.

22 Hölder's inequality, 2

Let E be a nonempty set, and let $1 \leq p, q \leq \infty$ be conjugate exponents. For each $g \in \ell^q(E)$, put

$$(22.1) \quad \lambda_g(f) = \sum_{x \in E} f(x) g(x)$$

when $f \in \ell^p(E)$. This makes sense, because of Hölder's inequality, and satisfies

$$(22.2) \quad |\lambda_g(f)| \leq \|f\|_p \|g\|_q.$$

Thus λ_g is a bounded linear functional on $\ell^p(E)$, with dual norm less than or equal to $\|g\|_q$. It is well known and not too difficult to show that the dual norm of λ_g on $\ell^p(E)$ is actually equal to $\|g\|_q$. If $q = 1$ and $p = \infty$, then one can also restrict λ_g to $c_0(E)$. One can also check that the dual norm of the restriction of λ_g to $c_0(E)$ with respect to the ℓ^∞ norm is also equal to $\|g\|_1$.

It is also well known that every bounded linear functional λ on $\ell^p(E)$ is of the form λ_g for some $g \in \ell^q(E)$ when $1 \leq p < \infty$, and that every bounded linear functional on $c_0(E)$ with respect to the ℓ^∞ norm is of the form λ_g for some $g \in \ell^1(E)$. The basic idea is to put

$$(22.3) \quad g(x) = \lambda(\delta_x),$$

where $\delta_x(x) = 1$ and $\delta_x(y) = 0$ when $y \in E$ and $y \neq x$. Using the boundedness of λ , one can show that $g \in \ell^q(E)$. By construction,

$$(22.4) \quad \lambda(f) = \lambda_g(f)$$

when $f(x) \neq 0$ for only finitely many $x \in E$. This implies the same relation for every $f \in \ell^p(E)$, $1 \leq p < \infty$, or $f \in c_0(E)$, as appropriate, because of the density of functions with finite support on E in these spaces.

If $0 < p < 1$, then $\ell^p(E) \subseteq \ell^1(E)$, and $\|f\|_1 \leq \|f\|_p$ for every $f \in \ell^p(E)$. It follows that the restriction of a bounded linear functional on $\ell^1(E)$ to $\ell^p(E)$ is a bounded linear functional with respect to the p -norm $\|f\|_p$. In particular, if $g \in \ell^\infty(E)$, then the restriction of λ_g to $\ell^p(E)$ is a bounded linear functional with dual norm less than or equal to $\|g\|_\infty$ with respect to $\|f\|_p$. One can check that the dual norm of λ_g on $\ell^p(E)$ is actually equal to $\|g\|_\infty$, because $\lambda_g(\delta_x) = g(x)$ and $\|\delta_x\|_p = 1$ for each $x \in E$.

Conversely, if λ is a bounded linear functional on $\ell^p(E)$, $0 < p < 1$, then $\lambda = \lambda_g$ for some $g \in \ell^\infty(E)$. The proof is basically the same as when $p = 1$. If g is as in (22.3), then g is bounded, and the ℓ^∞ norm of g is less than or equal to the dual norm of λ on $\ell^p(E)$, because $\|\delta_x\|_p = 1$ for each $x \in E$. One can then use density of functions with finite support in $\ell^p(E)$ to show that $\lambda = \lambda_g$.

23 Hilbert spaces

Let $(V, \langle v, w \rangle)$ be a real or complex inner product space, and put

$$(23.1) \quad \lambda_w(v) = \langle v, w \rangle$$

for each $w \in W$. By the Cauchy-Schwarz inequality, this is a bounded linear functional on V , with $\|\lambda_w\|_* \leq \|w\|$. More precisely,

$$(23.2) \quad \|\lambda\|_* = \|w\|,$$

because $\lambda(w) = \|w\|^2$. If V is complete, then it is well known that every bounded linear functional on V is of this form. Let us briefly review a proof of this fact.

Let $Y \subseteq V$, $Y \neq \emptyset$, and $z \in V$ be given, and let $\{y_j\}_{j=1}^\infty$ be a sequence of elements of Y such that

$$(23.3) \quad \lim_{j \rightarrow \infty} \|y_j - z\| = \inf\{\|y - z\| : y \in Y\}.$$

Note that

$$(23.4) \quad \left\| \frac{u+v}{2} \right\|^2 + \left\| \frac{u-v}{2} \right\|^2 = \frac{\|u\|^2}{2} + \frac{\|v\|^2}{2}$$

for every $u, v \in V$, which is a version of the *parallelogram law*. Applying this to $u = y_j - z$, $v = y_l - z$, we get that

$$(23.5) \quad \left\| \frac{y_j + y_l}{2} - z \right\|^2 + \frac{\|y_j - y_l\|^2}{4} = \frac{\|y_j - z\|^2}{2} + \frac{\|y_l - z\|^2}{2}$$

for each $j, l \geq 1$. If Y is convex, then $(y_j + y_l)/2 \in Y$ for every j, l , and hence

$$(23.6) \quad \inf\{\|y - z\| : y \in Y\} \leq \left\| \frac{y_j + y_l}{2} - z \right\|.$$

Combining this with (23.3) and (23.5), we get that

$$(23.7) \quad \lim_{j, l \rightarrow \infty} \|y_j - y_l\| = 0.$$

Thus $\{y_j\}_{j=1}^\infty$ is a Cauchy sequence when Y is convex. If V is complete and Y is also closed, then $\{y_j\}_{j=1}^\infty$ converges to an element y of Y with minimal distance to z .

If Y is a linear subspace of V , then one can show that $y \in Y$ has minimal distance to $z \in V$ if and only if $z - y$ is orthogonal to every element of Y . One can also check that y is uniquely determined by these properties. If V is complete, Y is a closed linear subspace of V , and $z \in V$, then it follows from that there is a $y \in Y$ such that $y - z$ is orthogonal to every element of Y .

Let λ be a bounded linear functional on V , and let

$$(23.8) \quad Y = \{v \in V : \lambda(v) = 0\}$$

be the kernel of λ . Thus Y is a closed linear subspace of V , and $Y = V$ if and only if $\lambda = 0$. If $\lambda \neq 0$, then there is a $w' \in V$ such that $w' \neq 0$ and $w' \perp y$ for every $y \in Y$, by the discussion in the previous paragraphs. In this case, one can check that $\lambda = \lambda_{w'}$, where w is a scalar multiple of w' . This uses the observation that Y has codimension 1 in V , so that every element of V can be expressed as a linear combination of w' and an element of Y .

24 The Hahn–Banach theorem

Let V be a real or complex vector space with a norm $\|v\|$, and let W be a linear subspace of V . The *Hahn–Banach theorem* states that every bounded linear functional on W can be extended to a bounded linear functional on V with the same norm. Note that this theorem does not work for p -norms, $0 < p < 1$. By standard arguments based on uniform continuity, a bounded linear functional on W has a unique extension to a bounded linear functional on the closure of W with the same norm, and this does work for p -norms on V .

It follows from the Hahn–Banach theorem that for every $v \in V$ with $v \neq 0$ there is a $\lambda \in V^*$ such that $\|\lambda\|_* = 1$ and

$$(24.1) \quad \lambda(v) = \|v\|.$$

More precisely, (24.1) determines a unique linear functional on the 1-dimensional subspace of V spanned by v , and the Hahn–Banach theorem implies that there is an extension of this linear functional to V with dual norm equal to 1. Note that this corollary does not hold for $\ell^p(E)$ when $0 < p < 1$ and E has at least two elements.

Let V be the space of continuous real or complex-valued functions f on the unit interval $[0, 1]$. If $0 < p < \infty$, then put

$$(24.2) \quad \|f\|_p = \left(\int_0^1 |f(x)|^p dx \right)^{1/p}.$$

One can check that this is a norm when $p \geq 1$ and a p -norm when $0 < p \leq 1$, in the same way as for ℓ^p . The counterpart of $\|f\|_p$ for $p = \infty$ is the supremum norm

$$(24.3) \quad \|f\|_\infty = \sup\{|f(x)| : 0 \leq x \leq 1\}.$$

It is well known that V is complete with respect to the supremum norm, and not with respect to $\|f\|_p$ when $0 < p < \infty$, for which the completions of V can be described in terms of Lebesgue integrals.

If $0 < p \leq q \leq \infty$, then

$$(24.4) \quad \|f\|_p \leq \|f\|_q$$

for every continuous function f on $[0, 1]$. This is easy to see when $q = \infty$, and it follows from the convexity of $r^{q/p}$ on the nonnegative real numbers when $q < \infty$. One can show that the only bounded linear functional on V with respect to $\|f\|_p$ is the trivial linear functional equal to 0 when $0 < p < 1$. This is because every continuous function f on $[0, 1]$ can be expressed as $\sum_{l=1}^n f_l$ for some continuous functions f_1, \dots, f_n such that $\sum_{l=1}^n \|f_l\|_p$ is arbitrarily small when $p < 1$. More precisely, one can choose the f_l 's to be supported on intervals of length approximately $1/n$.

25 Weak summability

Let E be a nonempty set, and let V be a real or complex vector space with a norm $\|v\|$. Also let $f(x)$ be a V -valued function on E such that $\sum_{x \in E} f(x)$

converges in the generalized sense. If λ is a bounded linear functional on V , then $\sum_{x \in E} \lambda(f(x))$ also converges in the generalized sense, and

$$(25.1) \quad \lambda\left(\sum_{x \in E} f(x)\right) = \sum_{x \in E} \lambda(f(x)).$$

Of course, $\sum_{x \in E} f(x)$ automatically converges in the generalized sense when $\|f(x)\|$ is summable on E , in which case $\lambda(f(x))$ is summable on E for every $\lambda \in V^*$, and

$$(25.2) \quad \sum_{x \in E} |\lambda(f(x))| \leq \|\lambda\|_* \sum_{x \in E} \|f(x)\|.$$

However, we have seen examples where $\sum_{x \in E} f(x)$ converges in the generalized sense, even though $\|f(x)\|$ is not summable on E . If $\phi(x)$ is a real or complex-valued function on E such that $\sum_{x \in E} \phi(x)$ converges in the generalized sense, then $\phi(x)$ is summable on E . In particular, $\lambda(f(x))$ is a summable function on E for every $\lambda \in V^*$ when $\sum_{x \in E} f(x)$ converges in the generalized sense.

26 Bounded partial sums

Let V be a real or complex vector space with a norm or p -norm $\|v\|$, $0 < p \leq 1$. Also let $X(V)$ be the space of sequences $\{v_j\}_{j=1}^\infty$ of elements of V such that the partial sums $\sum_{j=1}^n v_j$ of $\sum_{j=1}^\infty v_j$ are uniformly bounded in V . It is easy to see that $X(V)$ is a vector space with respect to termwise addition and scalar multiplication. Moreover,

$$(26.1) \quad \|\{v_j\}_{j=1}^\infty\|_{X(V)} = \sup_{n \geq 1} \left\| \sum_{j=1}^n v_j \right\|$$

is a norm or p -norm on $X(V)$, as appropriate. If $\{v_j\}_{j=1}^\infty \in X(V)$, then the sums $\sum_{j=l}^n v_j$ are uniformly bounded over $1 \leq l \leq n$, because

$$(26.2) \quad \sum_{j=l}^n v_j = \sum_{j=1}^n v_j - \sum_{j=1}^{l-1} v_j.$$

More precisely,

$$(26.3) \quad \left\| \sum_{j=l}^n v_j \right\| \leq \left\| \sum_{j=1}^n v_j \right\| + \left\| \sum_{j=1}^{l-1} v_j \right\| \leq 2 \|\{v_j\}_{j=1}^\infty\|_{X(V)}$$

when $\|v\|$ is a norm on V . Similarly,

$$(26.4) \quad \left\| \sum_{j=l}^n v_j \right\|^p \leq \left\| \sum_{j=1}^n v_j \right\|^p + \left\| \sum_{j=1}^{l-1} v_j \right\|^p \leq 2 \|\{v_j\}_{j=1}^\infty\|_{X(V)}^p$$

when $\|v\|$ is a p -norm on V , so that

$$(26.5) \quad \left\| \sum_{j=l}^n v_j \right\| \leq 2^{1/p} \|\{v_j\}_{j=1}^\infty\|_{X(V)}.$$

In particular, $\{v_j\}_{j=1}^\infty$ is bounded, by taking $l = n$.

An infinite series $\sum_{j=1}^\infty v_j$ with terms in V satisfies the ordinary Cauchy criterion if for every $\epsilon > 0$ there is an $L \geq 1$ such that

$$(26.6) \quad \left\| \sum_{j=l}^n v_j \right\| < \epsilon$$

when $n \geq l \geq L$. This is equivalent to saying that the sequence of partial sums $\sum_{j=1}^n v_j$ is a Cauchy sequence in V . Note that the partial sums are bounded in this case, so that $\{v_j\}_{j=1}^\infty \in X(V)$. Put

$$(26.7) \quad X_0(V) = \left\{ \{v_j\}_{j=1}^\infty \in X(V) : \sum_{j=1}^\infty v_j \text{ satisfies the Cauchy criterion} \right\}.$$

It is easy to see that $X_0(V)$ is a linear subspace of $X(V)$, and that $\{v_j\}_{j=1}^\infty$ is an element of $X_0(V)$ when $v_j = 0$ for all but finitely many j . One can also check that $X_0(V)$ is closed in $X(V)$, and in fact that $X_0(V)$ is the closure in $X(V)$ of the linear subspace of sequences $\{v_j\}_{j=1}^\infty$ such that $v_j = 0$ for all but finitely many j . If V is complete, then $X_0(V)$ is the same as the space of sequences $\{v_j\}_{j=1}^\infty$ such that $\sum_{j=1}^\infty v_j$ converges in V .

27 Bounded finite subsums

Let E be a nonempty set, and let V be a real or complex vector space with a norm or p -norm $\|v\|$, $0 < p \leq 1$. Also let $Y(E, V)$ be the space of V -valued functions $f(x)$ on E such that the sums $\sum_{x \in B} f(x)$ over nonempty finite subsets B of E are uniformly bounded in V . It is easy to see that this is a vector space with respect to pointwise addition and scalar multiplication, and that

$$(27.1) \quad \|f\|_{Y(E, V)} = \sup \left\{ \left\| \sum_{x \in B} f(x) \right\| : B \subseteq E, B \neq \emptyset, \text{ and } B \text{ has} \right. \\ \left. \text{only finitely many elements} \right\}$$

is a norm or p -norm on $Y(E, V)$, as appropriate. Note that each $f \in Y(E, V)$ is bounded, and that

$$(27.2) \quad \sup_{x \in E} \|f(x)\| \leq \|f\|_{Y(E, V)}.$$

Let $Y_0(E, V)$ be the set of V -valued functions $f(x)$ on E such that $\sum_{x \in E} f(x)$ satisfies the generalized Cauchy criterion. It is easy to see that this is a closed

linear subspace of $Y(E, V)$. If $f(x) = 0$ for all but finitely many $x \in E$, then $f \in Y_0(E, V)$, and in fact $Y_0(E, V)$ is the same as the closure in $Y(E, V)$ of the linear subspace of V -valued functions on E with finite support. If V is complete, then $Y_0(E, V)$ is also the same as the collection of V -valued functions $f(x)$ on E such that $\sum_{x \in E} f(x)$ converges in the generalized sense.

If $\|v\|$ is a norm on V and $\|f(x)\|$ is summable on E , or if $\|v\|$ is a p -norm on V and $\|f(x)\|$ is p -summable on E , $0 < p \leq 1$, then $f \in Y(E, V)$, and

$$(27.3) \quad \|f\|_{Y(E, V)}^p \leq \sum_{x \in E} \|f(x)\|^p.$$

Furthermore, $f \in Y_0(E, V)$ under these conditions. Conversely, if $V = \mathbf{R}$ and $f \in Y(E, \mathbf{R})$, then f is summable on E , and

$$(27.4) \quad \sum_{x \in E} |f(x)| \leq 2 \|f\|_{Y(E, \mathbf{R})}.$$

More precisely,

$$(27.5) \quad \sum_{\substack{x \in E \\ f(x) \geq 0}} f(x), \quad \sum_{\substack{x \in E \\ f(x) \leq 0}} -f(x) \leq \|f\|_{Y(E, \mathbf{R})}.$$

Similarly, if $V = \mathbf{C}$ and $f \in Y(E, \mathbf{C})$, then f is summable on E , and

$$(27.6) \quad \sum_{x \in E} |f(x)| \leq 4 \|f\|_{Y(E, \mathbf{C})}.$$

In this case, the real and imaginary parts $\operatorname{Re} f$, $\operatorname{Im} f$ of f are in $Y(E, \mathbf{R})$, and satisfy

$$(27.7) \quad \|\operatorname{Re} f\|_{Y(E, \mathbf{R})}, \|\operatorname{Im} f\|_{Y(E, \mathbf{R})} \leq \|f\|_{Y(E, \mathbf{C})}.$$

This implies the desired estimate for the ℓ^1 norm of f , which is less than or equal to the sum of the ℓ^1 norms of the real and imaginary parts of f .

28 Uniform boundedness

Let V be a real or complex vector space with a norm or p -norm $\|v\|$, and take $E = \mathbf{Z}_+$. Thus a V -valued function on E is basically the same as a sequence with terms in V , and $Y(\mathbf{Z}_+, V)$ can be identified with a linear subspace of $X(V)$. Also, $Y_0(\mathbf{Z}_+, V)$ corresponds to a linear subspace of $X_0(V)$ with respect to this identification, and the $X(V)$ norm is less than or equal to the $Y(\mathbf{Z}_+, V)$ norm. By definition, $Y(\mathbf{Z}_+, V)$, $Y_0(\mathbf{Z}_+, V)$, and the $Y(\mathbf{Z}_+, V)$ norm are invariant under one-to-one mappings of \mathbf{Z}_+ onto itself, while $X(V)$, $X_0(V)$, and the $X(V)$ norm are not invariant under rearrangements.

Suppose that $\{v_j\}_{j=1}^\infty$ is a sequence of elements of V such that $\{v_{\pi(j)}\}_{j=1}^\infty$ is an element of $X(V)$ for every one-to-one mapping π from \mathbf{Z}_+ onto itself, and let us show that $\{v_j\}_{j=1}^\infty$ corresponds to an element of $Y(\mathbf{Z}_+, V)$. This would be immediate if we also asked that the $X(V)$ norm of $\{v_{\pi(j)}\}_{j=1}^\infty$ be uniformly

bounded, independently of π . If $\{v_j\}_{j=1}^\infty$ does not correspond to an element of $Y(\mathbf{Z}_+, V)$, then there is a sequence of finite subsets B_1, B_2, \dots of E such that

$$(28.1) \quad \left\| \sum_{j \in B_n} v_j \right\| \rightarrow \infty \text{ as } n \rightarrow \infty.$$

One can also argue a bit more to get the B_n 's to be pairwise disjoint. This permits us to choose π so that $B_n = \{\pi(k_n), \dots, \pi(l_n)\}$ for some $k_n, l_n \in \mathbf{Z}_+$ with $k_n \leq l_n$ and every n . Hence $\{v_{\pi(j)}\}_{j=1}^\infty \notin X(V)$, as desired. Of course, the analogous statement for the generalized Cauchy criterion was discussed in Section 19.

29 Uniform boundedness, 2

Let M be a metric space, and let \mathcal{A} be a collection of continuous real or complex-valued functions on M . Suppose that \mathcal{A} is pointwise bounded on M , in the sense that

$$(29.1) \quad \mathcal{A}(x) = \{f(x) : f \in \mathcal{A}\}$$

is a bounded set in \mathbf{R} or \mathbf{C} , as appropriate, for each $x \in M$. Put

$$(29.2) \quad A_n = \{x \in M : |f(x)| \leq n \text{ for each } f \in \mathcal{A}\},$$

so that A_n is a closed set in M for each n , by continuity, and

$$(29.3) \quad \bigcup_{n=1}^{\infty} A_n = M,$$

by pointwise boundedness. If M is complete, then the Baire category theorem implies that A_n contains a nonempty open set in M for some n .

Suppose now that V is a real or complex vector space with a norm or p -norm, and that Λ is a collection of bounded linear functionals on V . If Λ is bounded pointwise on V and V is complete, then Λ is uniformly bounded on a nonempty open set in V , as in the previous paragraph. Using linearity, one can check that the elements of Λ have uniformly bounded dual norms. This is a version of the Banach–Steinhaus theorem, or uniform boundedness principle. Of course, Λ is uniformly bounded on bounded subsets of V when the dual norms of the elements of Λ are uniformly bounded.

Now let W be a real or complex vector space with a norm $\|w\|$, and let K be a subset of W . Suppose that

$$(29.4) \quad K(\lambda) = \{\lambda(w) : w \in K\}$$

is a bounded set in \mathbf{R} or \mathbf{C} , as appropriate, for each bounded linear functional λ on W . Each $w \in W$ determines a bounded linear functional on W^* , which sends $\lambda \in W^*$ to its value $\lambda(w)$ at w . Dual spaces are automatically complete, and so the boundedness of $K(\lambda)$ for each $\lambda \in W^*$ implies that the linear functionals $\lambda \mapsto \lambda(w)$ corresponding to $w \in K$ have uniformly bounded dual norm on W^* , as in the preceding paragraph. It follows that K is a bounded set in W , by the Hahn–Banach theorem.

30 Sums and linear functionals

Let E be a nonempty set, and let V be a real or complex vector space with a norm or p -norm $\|v\|$. If $f(x)$ is a V -valued function on E with uniformly bounded finite subsums, then $\lambda(f(x))$ has the same property for each bounded linear functional λ on V . Moreover,

$$(30.1) \quad \|\lambda \circ f\|_{Y(E, \mathbf{R})} \text{ or } \|\lambda \circ f\|_{Y(E, \mathbf{C})} \leq \|\lambda\|_* \|f\|_{Y(E, V)},$$

as appropriate. This implies that $\lambda(f(x))$ is summable on E , with

$$(30.2) \quad \sum_{x \in E} |\lambda(f(x))| \leq 2 \|\lambda\|_* \|f\|_{Y(E, V)}$$

in the real case, and

$$(30.3) \quad \sum_{x \in E} |\lambda(f(x))| \leq 4 \|\lambda\|_* \|f\|_{Y(E, V)}$$

in the complex case.

Conversely,

$$(30.4) \quad \left| \lambda \left(\sum_{x \in B} f(x) \right) \right| = \left| \sum_{x \in B} \lambda(f(x)) \right| \leq \sum_{x \in B} |\lambda(f(x))|$$

for every finite set $B \subseteq E$ and $\lambda \in V^*$. Suppose that $\lambda(f(x))$ is summable on E for each $\lambda \in V^*$, and that

$$(30.5) \quad \sum_{x \in E} |\lambda(f(x))| \leq C \|\lambda\|_*$$

for some $C \geq 0$ and every $\lambda \in V^*$. If $\|v\|$ is a norm on V , then the Hahn–Banach theorem implies that

$$(30.6) \quad \left\| \sum_{x \in B} f(x) \right\| \leq C$$

for every finite set $B \subseteq E$. Hence $f \in Y(E, V)$ and

$$(30.7) \quad \|f\|_{Y(E, V)} \leq C$$

under these conditions.

Let K be the set of vectors in V of the form $\sum_{x \in B} f(x)$, where $B \subseteq E$ is a finite set. If $\lambda(f(x))$ is summable on E for some $\lambda \in V^*$, then the set $K(\lambda)$ as in (29.4) is bounded. If $\lambda(f(x))$ is summable on E for every $\lambda \in V^*$, and if $\|v\|$ is a norm on V , then it follows that K is a bounded set in V , as in the previous section. This is the same as saying that $f \in Y(E, V)$.

31 Seminorms

Let V be a vector space over the real or complex numbers. A nonnegative real-valued function $N(v)$ on V is said to be a *seminorm* if

$$(31.1) \quad N(tv) = |t|N(v)$$

for every $v \in V$ and $t \in \mathbf{R}$ or \mathbf{C} , as appropriate, and

$$(31.2) \quad N(v+w) \leq N(v) + N(w)$$

for every $v, w \in V$. Thus a seminorm $N(v)$ is a norm exactly when $N(v) > 0$ for every $v \in V$ with $v \neq 0$. As another class of examples, $N_\lambda(v) = |\lambda(v)|$ is a seminorm on V when λ is a linear functional on V . Observe that

$$(31.3) \quad \{v \in V : N(v) = 0\}$$

is a linear subspace of V when $N(v)$ is a seminorm on V .

Let \mathcal{N} be a collection of seminorms on V . Let us say that $U \subseteq V$ is an open set with respect to \mathcal{N} if for every $u \in U$ there are finitely many seminorms $N_1, \dots, N_l \in \mathcal{N}$ and positive real numbers r_1, \dots, r_l such that

$$(31.4) \quad \{v \in V : N_j(u-v) < r_j, j = 1, \dots, l\} \subseteq U.$$

It is easy to see that this defines a topology on V . Note that this topology is Hausdorff if and only if \mathcal{N} satisfies the positivity condition that for each $v \in V$ with $v \neq 0$ there is an $N \in \mathcal{N}$ such that $N(v) > 0$. If \mathcal{N} consists of a single norm, then this is the usual topology associated to the norm.

Suppose that V is equipped with a norm or p -norm $\|v\|_V$, and consider the collection of seminorms on V of the form $N_\lambda(v) = |\lambda(v)|$, where $\lambda \in V^*$. The topology on V associated to this collection of seminorms is known as the *weak topology*. If $\|v\|_V$ is a norm on V , then the Hahn–Banach theorem implies that for each $v \in V$ with $v \neq 0$ there is a $\lambda \in V^*$ such that $\lambda(v) \neq 0$. Thus $N_\lambda(v) > 0$, and so the weak topology on V is Hausdorff when $\|v\|_V$ is a norm. Note that open subsets of V with respect to the weak topology are open with respect to $\|v\|_V$, because the linear functionals being used are bounded.

Now let W be a real or complex vector space with a norm or p -norm $\|w\|_W$, and consider $V = W^*$. Each $w \in W$ determines a linear functional $\lambda \mapsto \lambda(w)$ on W^* , and hence a seminorm $N_w^*(\lambda) = |\lambda(w)|$ on W^* . The topology on W^* defined by this collection of seminorms is known as the *weak* topology*. This topology is automatically Hausdorff, but it is helpful for $\|w\|_W$ to be a norm on W so that there are plenty of bounded linear functionals on W . Note that every open set in W^* with respect to the weak* topology is also open with respect to the dual norm on W^* .

32 Sums in dual spaces

Let E be a nonempty set, let W be a real or complex vector space with a norm or p -norm $\|w\|$, and let f be a function on E with values in the dual W^* of W .

Suppose that $f(x)(w)$ is a summable function on E for every $w \in W$, where $f(x)(w)$ refers to the value of $f(x) \in W^*$ at w , and that

$$(32.1) \quad \sum_{x \in E} |f(x)(w)| \leq C \|w\|$$

for some $C \geq 0$ and every $w \in W$. In this case, $\sum_{x \in E} f(x)(w)$ defines a bounded linear functional on W with dual norm $\leq C$. One can also say that $\sum_{x \in E} f(x)$ converges in the generalized sense with respect to the weak* topology on W^* under these conditions.

This estimate also implies that

$$(32.2) \quad \left| \sum_{x \in B} f(x)(w) \right| \leq C \|w\|$$

for every finite set $B \subseteq E$ and $w \in W$, which is to say that

$$(32.3) \quad \left\| \sum_{x \in B} f(x) \right\|_* \leq C$$

for every finite set $B \subseteq E$. Thus $f \in Y(E, W^*)$, and

$$(32.4) \quad \|f\|_{Y(E, W^*)} \leq C.$$

Conversely, if $f \in Y(E, W^*)$, then $f(x)(w)$ is summable on E for every $w \in W$, with ℓ^1 norm bounded by $2 \|f\|_{Y(E, V)}$ in the real case and by $4 \|f\|_{Y(E, V)}$ in the complex case. If W is complete and $f(x)(w)$ is summable on E for every $w \in W$, then one can use the uniform boundedness principle to conclude that $f \in Y(E, W^*)$.

33 Seminorms, 2

Let V be a vector space over the real or complex numbers, and let N_1, N_2, \dots be a sequence of seminorms on V such that for each $v \in V$ with $v \neq 0$ there is a positive integer j for which $N_j(v) > 0$. Under these conditions, one can check that

$$(33.1) \quad d(v, w) = \max\{\min(N_j(v - w), 1/j) : j \in \mathbf{Z}_+\}$$

defines a metric on V , and that the topology on V determined by this metric is the same as the one associated to this sequence of seminorms as in Section 31. Conversely, if the topology on V determined by a collection \mathcal{N} of seminorms on V is metrizable, then it is Hausdorff, and there is a countable local base for the topology at 0. Using the latter, one can show that there is a subcollection of \mathcal{N} with only finitely or countably many elements that determines the same topology on V .

Suppose now that V is equipped with a norm or p -norm $\|v\|$, and consider the weak topology on V . Suppose also that for each $v \in V$ with $v \neq 0$ there

is a $\lambda \in V^*$ such that $\lambda(v) \neq 0$, which follows from the Hahn–Banach theorem when $\|v\|$ is a norm on V , and which implies that the weak topology on V is Hausdorff. Suppose in addition that V^* is separable, and let $\lambda_1, \lambda_2, \dots$ be a sequence of bounded linear functionals on V whose linear span is dense in V^* . Let $N_{\lambda_1}, N_{\lambda_2}, \dots$ be the seminorms on V corresponding to the λ_j 's as in Section 31. Under these conditions, one can check that the topology induced on a bounded set in V by the weak topology is the same as the topology induced by the seminorms $N_{\lambda_1}, N_{\lambda_2}, \dots$, and hence is metrizable.

Similarly, we can consider the weak* topology on the dual of a vector space V with a norm or p -norm. Suppose that V is separable, so that there is a sequence of vectors v_1, v_2, \dots in V whose linear span is dense in V . Let $N_{v_1}^*, N_{v_2}^*, \dots$ be the seminorms on V^* corresponding to the v_j 's, as in Section 31. If K is a bounded set in V^* with respect to the dual norm, then one can again check that the topology induced on K by the weak* topology is the same as the topology induced by the seminorms $N_{v_1}^*, N_{v_2}^*, \dots$, and is therefore metrizable.

Note that the unit ball

$$(33.2) \quad B_1^* = \{\lambda \in V^* : \|\lambda\|_* \leq 1\}$$

in the dual V^* of V is closed with respect to the weak* topology. To see this, it is convenient to describe B_1^* as the set of $\lambda \in V^*$ such that

$$(33.3) \quad |\lambda(v)| \leq 1$$

for every $v \in V$ with $\|v\| \leq 1$. The Banach–Alaoglu theorem states that B_1^* is actually compact with respect to the weak* topology. If V is separable, then the topology induced on B_1^* by the weak* topology on V^* is metrizable, as in the previous paragraph. In this case, compactness of B_1^* in the weak* topology is equivalent to sequential compactness.

34 Isometric embeddings

Let $(M, d(x, y))$ be a metric space. It is easy to check that

$$(34.1) \quad f_p(x) = d(p, x)$$

is a continuous function on M for each $p \in M$, using the triangle inequality. If M is bounded, then f_p is also a bounded function on M . Thus $p \mapsto f_p$ defines a mapping from M into the space $C_b(M)$ of bounded continuous real-valued functions on M . Using the triangle inequality, one can show that this is an isometric embedding of M into $C_b(M)$ with the supremum norm.

If M is not bounded, then one can pick a basepoint $p_0 \in M$, and put

$$(34.2) \quad \tilde{f}_p = f_p - f_{p_0}.$$

Using the triangle inequality again, one can check that \tilde{f}_p is a bounded function on M for each $p \in M$. Moreover, $p \mapsto \tilde{f}_p$ is an isometric embedding of M into

$C_b(M)$ for the same reasons as before, since

$$(34.3) \quad \tilde{f}_p - \tilde{f}_q = f_p - f_q$$

for every $p, q \in M$.

Suppose now that V is a real or complex vector space with a norm $\|v\|$, and let B_1^* be the closed unit ball in the dual space V^* , as in (33.2). Each $v \in V$ determines a bounded linear functional on V^* defined by

$$(34.4) \quad L_v(\lambda) = \lambda(v),$$

which can also be considered as a bounded continuous function on B_1^* with respect to the topology induced by the weak* topology. Thus $v \mapsto L_v$ defines a linear mapping from V into the space $C(B_1^*)$ of continuous real or complex-valued functions on B_1^* with respect to the weak* topology, as appropriate. By the Banach–Alaoglu theorem, B_1^* is a compact Hausdorff space with respect to this topology. Using the Hahn–Banach theorem, it is easy to see that $v \mapsto L_v$ is also an isometry from V into $C(B_1^*)$, with respect to the supremum norm on $C(B_1^*)$.

Part II

Functions, measures, and paths

35 Uniform boundedness, 3

Let (X, \mathcal{A}) be a measurable space, which is to say a set X with a σ -algebra \mathcal{A} of measurable subsets of X , and let p be a nonnegative real-valued function on \mathcal{A} . Suppose that for every sequence A_1, A_2, \dots of pairwise-disjoint measurable subsets of X ,

$$(35.1) \quad p\left(\bigcup_{j=1}^{\infty} A_j\right) \leq \sum_{j=1}^{\infty} p(A_j) < \infty.$$

This implies that $p(\emptyset) = 0$, by taking $A_j = \emptyset$ for each j .

Let B_1, B_2, \dots be a decreasing sequence of measurable subsets of X , so that $B_{j+1} \subseteq B_j$ for each j , and put $B_{\infty} = \bigcap_{j=1}^{\infty} B_j$. Thus $A_j = B_j \setminus B_{j+1}$ is a sequence of pairwise-disjoint measurable subsets of X which are also disjoint from B_{∞} , and

$$(35.2) \quad B_n = \left(\bigcup_{j=n}^{\infty} A_j\right) \cup B_{\infty}$$

for each n . In particular, $\sum_{j=1}^{\infty} p(A_j)$ converges, which implies that $p(B_n)$ is uniformly bounded in n , since

$$(35.3) \quad p(B_n) \leq \sum_{j=n}^{\infty} p(A_j) + p(B_{\infty})$$

for each n . If $B_\infty = \emptyset$, then (35.3) implies that $\{p(B_n)\}_{n=1}^\infty$ converges to 0. If C_1, C_2, \dots is an increasing sequence of measurable subsets of X , then a similar argument shows that $p(C_n)$ is uniformly bounded in n , but we shall not need this here.

If $A \subseteq X$ is measurable, then put

$$(35.4) \quad p^*(A) = \sup \left\{ \sum_{j=1}^{\infty} p(A_j) : A_1, A_2, \dots \text{ are pairwise-disjoint} \right. \\ \left. \text{measurable subsets of } X \text{ such that } A = \bigcup_{j=1}^{\infty} A_j \right\}.$$

We would like to show that $p^*(A) < \infty$ under these conditions. Equivalently, one can check that

$$(35.5) \quad p^*(A) = \sup \left\{ \sum_{j=1}^n p(A_j) : A_1, \dots, A_n \text{ are pairwise-disjoint} \right. \\ \left. \text{measurable subsets of } X \text{ such that } A = \bigcup_{j=1}^n A_j \right\}.$$

More precisely, the second definition of $p^*(A)$ is clearly less than or equal to the first definition, because a partition of A into finitely many measurable sets can be extended to an infinite partition using the empty set. To show that the first definition of $p^*(A)$ is less than or equal to the second definition, one can approximate an infinite partition A_1, A_2, \dots of A by the finite partitions consisting of the sets A_1, \dots, A_n and $\bigcup_{j=n+1}^{\infty} A_j$ for each n .

If B_1, B_2, \dots is a sequence of pairwise-disjoint measurable subsets of X , then

$$(35.6) \quad \sum_{l=1}^{\infty} p^*(B_l) \leq p^*\left(\bigcup_{l=1}^{\infty} B_l\right),$$

because partitions of the B_l 's can be combined to get a partition of $\bigcup_{l=1}^{\infty} B_l$. Similarly,

$$(35.7) \quad p^*\left(\bigcup_{l=1}^{\infty} B_l\right) \leq \sum_{l=1}^{\infty} p^*(B_l),$$

because every measurable partition $\{E_j\}_{j=1}^\infty$ of $\bigcup_{l=1}^\infty B_l$ can be refined to get a partition $\{E_j \cap B_l\}_{j,l=1}^\infty$ which is a combination of partitions of the B_l 's. Countable subadditivity implies that $p(E_j)$ is less than or equal to the sum of $p(E_j \cap B_l)$ over l for each j , so that the sum of $p(E_j)$ over j is less than or equal to the sum of $p(E_j \cap B_l)$ over j and l . The sum of $p(E_j \cap B_l)$ over j is less than or equal to $p^*(B_l)$ for each l , and so the sum of $p(E_j \cap B_l)$ over j and l is less than or equal to the sum of $p^*(B_l)$ over l , as desired. Therefore

$$(35.8) \quad p^*\left(\bigcup_{l=1}^{\infty} B_l\right) = \sum_{l=1}^{\infty} p^*(B_l),$$

which means that p^* is countably additive.

Suppose for the sake of a contradiction that $p^*(A) = \infty$ for some measurable set $A \subseteq X$. This implies that there is a finite sequence of pairwise-disjoint measurable subsets $A_{1,1}, \dots, A_{1,n_1}$ of X such that

$$(35.9) \quad A = \bigcup_{j=1}^{n_1} A_{1,j}$$

and

$$(35.10) \quad \sum_{j=1}^{n_1} p(A_{1,j}) \geq 1.$$

We also have that $p^*(A_{1,j}) = \infty$ for some j , since

$$(35.11) \quad p^*(A) = p^*(A_{1,1}) + \dots + p^*(A_{1,n_1}),$$

and so we can relabel the indices, if necessary, to get that

$$(35.12) \quad p^*(A_{1,n_1}) = \infty.$$

This permits us to repeat the process, to get a finite sequence $A_{2,1}, \dots, A_{2,n_2}$ of pairwise-disjoint measurable subsets of X such that

$$(35.13) \quad A_{1,n_1} = \sum_{j=1}^{n_2} A_{2,j}$$

and

$$(35.14) \quad \sum_{j=1}^{n_2} p(A_{2,j}) \geq 2.$$

As before, $p^*(A_{2,j}) = \infty$ for some j , and we can relabel the indices if necessary to get that $p^*(A_{2,n_2}) = \infty$. Continuing in this way, we get a finite sequence $A_{k,1}, \dots, A_{k,n_k}$ of pairwise-disjoint measurable subsets of X for each positive integer k such that

$$(35.15) \quad \bigcup_{l=1}^{n_k} A_{k,l} = A_{k-1,n_{k-1}}$$

when $k \geq 2$,

$$(35.16) \quad \sum_{l=1}^{n_k} p(A_{k,l}) \geq k,$$

and $p^*(A_{k,n_k}) = \infty$.

However,

$$(35.17) \quad \sum_{k=1}^{\infty} \sum_{l=1}^{n_k-1} p(A_{k,l}) < \infty,$$

because the $A_{k,l}$'s are pairwise disjoint when $l < n_k$. Hence the sums

$$(35.18) \quad \sum_{l=1}^{n_k-1} p(A_{k,l})$$

are uniformly bounded in k , and even converge to 0 as $k \rightarrow \infty$. By construction, $A_{k+1,n_{k+1}} \subseteq A_{k,n_k}$ for each k , and so $p(A_{k,n_k})$ is also uniformly bounded in k , as mentioned earlier in the section. This implies that the sums

$$(35.19) \quad \sum_{l=1}^{n_k} p(A_{k,l}) = \sum_{l=1}^{n_k-1} p(A_{k,l}) + p(A_{k,n_k})$$

are uniformly bounded in k as well. This contradicts (35.16), and we conclude that $p^*(A) < \infty$ for every measurable set $A \subseteq X$.

Of course,

$$(35.20) \quad p(A) \leq p^*(A)$$

for every measurable set $A \subseteq X$, and in fact p^* is the smallest countably-additive measure with this property. More precisely, if ρ is a countably-additive measure such that $p(A) \leq \rho(A)$ for every measurable set $A \subseteq X$, then $p^*(A) \leq \rho(A)$ for each A . This follows directly from the definition of $p^*(A)$. Observe too that the hypothesis that $\sum_{j=1}^{\infty} p(A_j)$ converges when A_1, A_2, \dots is a sequence of pairwise-disjoint measurable sets is necessary in order to have a finite measure ρ such that $p(A) \leq \rho(A)$.

36 Real and complex measures

Let (X, \mathcal{A}) be a measurable space, and let μ be a real or complex measure on this space. This means that μ is a real or complex-valued function on \mathcal{A} such that

$$(36.1) \quad \mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu(A_j)$$

for every sequence A_1, A_2, \dots of pairwise-disjoint measurable subsets of X . More precisely, the convergence of the series $\sum_{j=1}^{\infty} \mu(A_j)$ is part of the definition. It follows that the series converges absolutely, because every rearrangement of the series is of the same type. Note that $\mu(\emptyset) = 0$ is also implied by the definition, by taking $A_j = \emptyset$. If $p(A) = |\mu(A)|$, then it is easy to see that $p(A)$ satisfies the conditions described in the previous section. Hence $p^*(A)$ is a countably-additive finite measure, which is commonly denoted $|\mu|(A)$.

In the real case, μ is also known as a signed measure on X , and it is easy to see that

$$(36.2) \quad \mu^+(A) = \frac{|\mu|(A) + \mu(A)}{2}, \quad \mu^-(A) = \frac{|\mu|(A) - \mu(A)}{2}$$

are finite nonnegative measures on X . Note that

$$(36.3) \quad \mu(A) = \mu^+(A) - \mu^-(A)$$

and

$$(36.4) \quad |\mu|(A) = \mu^+(A) + \mu^-(A)$$

for each measurable set $A \subseteq X$. Similarly, if μ is a complex measure on X , then μ can be expressed as a linear combination of finite nonnegative measures on X , by applying this argument to the real and imaginary parts of μ .

There are a number of simplifications that can be made in the previous section when $p(A) = |\mu(A)|$ for a real measure μ on X . The first simplification is to replace the earlier definition of $p^*(A)$ with

$$(36.5) \quad p^*(A) = \sup\{|\mu(B)| + |\mu(C)| : B, C \in \mathcal{A}, A = B \cup C, B \cap C = \emptyset\}.$$

The right side is clearly less than or equal to the earlier definition of $p^*(A)$. To show the opposite inequality, let $\{A_j\}_{j=1}^\infty$ be any sequence of pairwise-disjoint measurable subsets of X such that $A = \bigcup_{j=1}^\infty A_j$. If B is the union of the A_j 's with $\mu(A_j) \geq 0$ and C is the union of the A_j 's with $\mu(A_j) < 0$, then $A = B \cup C$, $B \cap C = \emptyset$, and

$$(36.6) \quad \sum_{j=1}^\infty |\mu(A_j)| = \mu(B) - \mu(C) = |\mu(B)| + |\mu(C)|.$$

This implies that the earlier definition of $p^*(A)$ is less than or equal to the right side of (36.5), by taking the supremum over all such sequences $\{A_j\}_{j=1}^\infty$. In the same way, we also have that

$$(36.7) \quad p^*(A) = \sup\{\mu(B) - \mu(C) : B, C \in \mathcal{A}, A = B \cup C, B \cap C = \emptyset\}.$$

This makes it much easier to show that $p^*(A) < \infty$. If $p^*(A) = \infty$ for some measurable set $A \subseteq X$, then there are disjoint measurable sets B, C such that $A = B \cup C$ and $\mu(B) - \mu(C)$ is as large as we want. Of course,

$$(36.8) \quad \mu(A) = \mu(B) + \mu(C),$$

which implies that both $|\mu(B)|$ and $|\mu(C)|$ are as large as we want. Because p^* is subadditive, we also have that $p^*(B) = \infty$ or $p^*(C) = \infty$. Put $A_1 = B$ if $p^*(B) = \infty$, and otherwise $A_1 = C$. Repeating the process, we get a decreasing sequence $\{A_j\}_{j=1}^\infty$ of measurable subsets of X such that $p^*(A_l) = \infty$ for each l and $|\mu(A_l)| \rightarrow \infty$ as $l \rightarrow \infty$. This contradicts the fact that $|\mu(A_l)|$ is bounded when $A_{l+1} \subseteq A_l$ for each l , as in the previous section. One can also use the fact that $\{\mu(A_j)\}_{j=1}^\infty$ converges under these conditions, and hence is bounded, which is based on a similar argument. It follows that $p^*(A) < \infty$ when $p(A) = |\mu(A)|$ for a complex measure μ , by considering the real and imaginary parts of μ .

In the real case, we can combine (36.7) and (36.8) to get that

$$(36.9) \quad \mu^+(A) = \sup\{\mu(B) : B \in \mathcal{A}, B \subseteq A\}.$$

We may restrict our attention to $B \subseteq A$ such that $\mu(B) \geq 0$ here, since $B = \emptyset$ has these properties. Similarly,

$$(36.10) \quad \mu^-(A) = \sup\{-\mu(C) : C \in \mathcal{A}, C \subseteq A\}.$$

If μ_1 is a nonnegative real measure on X such that $\mu(A) \leq \mu_1(A)$ for every measurable set $A \subseteq X$, then

$$(36.11) \quad \mu^+(A) \leq \mu_1(A)$$

for every $A \in \mathcal{A}$. More precisely, this uses the fact that

$$(36.12) \quad \mu_1(A) = \mu_1(B) + \mu_1(A \setminus B) \geq \mu_1(B)$$

when $B \subseteq A$, because $\mu_1(A \setminus B) \geq 0$. Similarly, if μ_2 is a nonnegative real measure on X such that $\mu(A) \geq -\mu_2(A)$ for every measurable set $A \subseteq X$, then

$$(36.13) \quad \mu^-(A) \leq \mu_2(A)$$

for every $A \in \mathcal{A}$. Of course, $\mu_1 = \mu^+$ and $\mu_2 = \mu^-$ have these properties, by construction.

If μ_1, μ_2 are finite nonnegative real measures on X such that

$$(36.14) \quad \mu(A) = \mu_1(A) - \mu_2(A)$$

for every measurable set $A \subseteq X$, then

$$(36.15) \quad -\mu_2(A) \leq \mu(A) \leq \mu_1(A)$$

for every $A \in \mathcal{A}$. Thus μ_1 and μ_2 satisfy (36.11) and (36.13), respectively, as in the preceding paragraph. As before, $\mu_1 = \mu^+$, $\mu_2 = \mu^-$ have this property, by construction.

Suppose that P, Q are disjoint measurable subsets of X such that $P \cup Q = X$ and

$$(36.16) \quad |\mu|(X) = \mu(P) - \mu(Q).$$

This is the same as saying that the supremum in (36.7) is attained when $A = X$, with $B = P$ and $C = Q$. If E is a measurable subset of P such that $\mu(E) < 0$, then

$$(36.17) \quad \mu(P) = \mu(P \setminus E) + \mu(E) < \mu(P \setminus E)$$

and

$$(36.18) \quad \mu(Q) > \mu(Q) + \mu(E) = \mu(Q \cup E),$$

which implies that

$$(36.19) \quad \mu(P) - \mu(Q) < \mu(P \setminus E) - \mu(Q \cup E),$$

contradicting maximality. Thus $\mu(E) \geq 0$ for every measurable set $E \subseteq P$, and similarly $\mu(E) \leq 0$ for every measurable set $E \subseteq Q$. Using this, one can check that

$$(36.20) \quad \mu^+(A) = \mu(A \cap P), \quad \mu^-(A) = \mu(A \cap Q)$$

for every measurable set $A \subseteq X$, which is to say that the suprema in (36.9) and (36.10) are attained with $B = A \cap P$ and $C = A \cap Q$.

The *Hahn decomposition theorem* states that there are disjoint measurable subsets P, Q of X such that $P \cup Q = X$ and (36.20) holds for every measurable set $A \subseteq X$. One way to prove this is to show that the supremum in (36.7) is attained when $A = X$, as in the next paragraph. Another way is to use the Radon–Nikodym theorem, discussed in Section 38.

Suppose that $\{B_j\}_{j=1}^\infty, \{C_j\}_{j=1}^\infty$ are sequences of measurable subsets of X such that $B_j \cap C_j = \emptyset$ and $B_j \cup C_j = X$ for each j , and

$$(36.21) \quad \lim_{j \rightarrow \infty} (\mu(B_j) - \mu(C_j)) = |\mu|(X).$$

Observe that

$$(36.22) \quad |\mu|(X) - (\mu(B_j) - \mu(C_j)) = 2(\mu^-(B_j) + \mu^+(C_j))$$

for each j , because $|\mu|(X) = |\mu|(B_j) + |\mu|(C_j)$. Hence

$$(36.23) \quad \lim_{j \rightarrow \infty} \mu^-(B_j) = \lim_{j \rightarrow \infty} \mu^+(C_j) = 0.$$

Using this, one can show that $\{B_j\}_{j=1}^\infty, \{C_j\}_{j=1}^\infty$ are Cauchy sequences with respect to the semimetric on \mathcal{A} associated to $|\mu|$ as in Section 79, and hence converge. This is equivalent to saying that the sequences of their indicator functions are Cauchy sequences in $L^1(X, |\mu|)$, and hence converge in $L^1(X, |\mu|)$ to indicator functions of measurable subsets of X . More precisely, (36.23) implies that $\{B_j\}_{j=1}^\infty$ converges to the empty set with respect to μ^- , and that $\{C_j\}_{j=1}^\infty$ converges to the empty set with respect to μ^+ . This implies in turn that $\{B_j\}_{j=1}^\infty$ converges to X with respect to μ^+ , and that $\{C_j\}_{j=1}^\infty$ converges to X with respect to μ^- , because $C_j = X \setminus B_j$ for each j . It follows that $\{B_j\}_{j=1}^\infty, \{C_j\}_{j=1}^\infty$ are Cauchy sequences with respect to both μ^+ and μ^- , and are thus Cauchy sequences with respect to $|\mu| = \mu^+ + \mu^-$. The limits of these sequences correspond to measurable subsets P, Q of X that are determined up to sets of $|\mu|$ -measure 0. By construction, Q is the same as $X \setminus P$ up to a set of $|\mu|$ -measure 0, and we may as well take $Q = X \setminus P$. We also have that $\mu^-(P) = \mu^+(Q) = 0$, $|\mu|(X) = \mu(P) - \mu(Q)$, and so on.

37 Vector-valued measures

Let (X, \mathcal{A}) be a measurable space, and let V be a real or complex vector space with a norm $\|v\|$. More precisely, suppose that V is a Banach space, which means that V is complete with respect to the metric associated to the norm. Let μ be a V -valued function on \mathcal{A} such that

$$(37.1) \quad \mu\left(\bigcup_{j=1}^\infty A_j\right) = \sum_{j=1}^\infty \mu(A_j)$$

for every sequence A_1, A_2, \dots of pairwise-disjoint measurable subsets of X . Again convergence of the sum

$$(37.2) \quad \sum_{j=1}^\infty \mu(A_j)$$

is part of the hypothesis, which implies convergence of rearrangements of the sum. However, in this case, absolute convergence

$$(37.3) \quad \sum_{j=1}^{\infty} \|\mu(A_j)\| < \infty$$

is an additional condition. If we have absolute convergence, then $p(A) = \|\mu(A)\|$ satisfies the requirements of Section 35. This implies that $\|\mu\|(A) = p^*(A)$ is a countably-additive finite nonnegative measure.

Let ν be a countably-additive finite nonnegative measure on (X, \mathcal{A}) , and take V to be $L^q(X, \nu)$ for some q , $1 \leq q < \infty$. Also let $\mathbf{1}_A(x)$ be the indicator function of $A \subseteq X$, equal to 1 when $x \in A$ and to 0 when $x \in X \setminus A$. If $\mu(A) = \mathbf{1}_A$ for each measurable set $A \subseteq X$, then μ is a V -valued function on \mathcal{A} that satisfies the countable additivity condition described in the previous paragraph. If $q = 1$, then μ also satisfies the absolute convergence condition. This does not normally work when $q > 1$, even when ν is Lebesgue measure on the unit interval.

Let μ be an arbitrary V -valued function μ on \mathcal{A} that satisfies the countable additivity condition mentioned at the beginning of the section, not necessarily with absolute convergence. If λ is a bounded linear functional on V , then

$$(37.4) \quad \mu_\lambda(A) = \lambda(\mu(A))$$

defines a real or complex measure on (X, \mathcal{A}) , as appropriate. In particular, μ_λ has finite total variation $|\mu_\lambda|$, and

$$(37.5) \quad |\mu_\lambda(A)| \leq |\mu_\lambda|(A) \leq |\mu_\lambda|(X)$$

for every measurable set $A \subseteq X$. Thus

$$(37.6) \quad \{\lambda(\mu(A)) : A \in \mathcal{A}\}$$

is a bounded set of real or complex numbers, as appropriate, for each $\lambda \in V^*$. It follows that

$$(37.7) \quad \{\mu(A) : A \in \mathcal{A}\}$$

is a bounded set in V , as in Section 29.

If α is a real measure on (X, \mathcal{A}) , then

$$(37.8) \quad |\alpha|(X) \leq 2 \sup\{|\alpha(A)| : A \in \mathcal{A}\},$$

because of (36.5). Similarly, if β is a complex measure on (X, \mathcal{A}) , then

$$(37.9) \quad |\beta|(X) \leq 4 \sup\{|\beta(A)| : A \in \mathcal{A}\},$$

by applying (37.8) to the real and imaginary parts of β . If μ is a countably-additive V -valued function on \mathcal{A} and λ is a bounded linear functional on V , as in the previous paragraph, then

$$(37.10) \quad |\mu_\lambda(A)| = |\lambda(\mu(A))| \leq \|\lambda\|_* \|\mu(A)\|$$

for every measurable set $A \subseteq X$. Hence

$$(37.11) \quad |\mu_\lambda|(X) \leq 2 \|\lambda\|_* \sup\{\|\mu(A)\| : A \in \mathcal{A}\}$$

in the real case, and

$$(37.12) \quad |\mu_\lambda|(X) \leq 4 \|\lambda\|_* \sup\{\|\mu(A)\| : A \in \mathcal{A}\}$$

in the complex case.

If μ is a countably-additive V -valued function on \mathcal{A} and B_1, B_2, \dots is an increasing sequence of measurable subsets of X , then

$$(37.13) \quad \lim_{j \rightarrow \infty} \mu(B_j) = \mu\left(\bigcup_{j=1}^{\infty} B_j\right).$$

This follows from countable additivity by taking $A_1 = B_1$ and $A_j = B_j \setminus B_{j-1}$ when $j \geq 2$, as usual. Conversely, this continuity condition implies countable additivity when μ is finitely additive, by taking $B_n = \bigcup_{j=1}^n A_j$. Similarly, if C_1, C_2, \dots is a decreasing sequence of measurable subsets of X , then

$$(37.14) \quad \lim_{l \rightarrow \infty} \mu(C_l) = \mu\left(\bigcap_{l=1}^{\infty} C_l\right).$$

This is equivalent to (37.13) when μ is finitely additive, with $B_j = X \setminus C_j$.

Let us use these continuity conditions to give another proof of the fact that μ is bounded, like the one for real measures in the previous section. Put

$$(37.15) \quad \widehat{\mu}(A) = \sup\{\|\mu(B)\| : B \in \mathcal{A}, B \subseteq A\}$$

for each measurable set $A \subseteq X$, which may be $+\infty$ a priori. Observe that

$$(37.16) \quad \widehat{\mu}(A \cup A') \leq \widehat{\mu}(A) + \widehat{\mu}(A')$$

for any measurable sets $A, A' \subseteq X$. This is because any measurable subset B of $A \cup A'$ can be expressed as the union of $B \cap A \subseteq A$ and $B \setminus A \subseteq A'$, which are automatically disjoint. Thus $\mu(B)$ is the sum of $\mu(B \cap A)$ and $\mu(B \setminus A)$, so that $\|\mu(B)\|$ is less than or equal to the sum of $\|\mu(B \cap A)\|$ and $\|\mu(B \setminus A)\|$, which is less than or equal to the sum of $\widehat{\mu}(A)$ and $\widehat{\mu}(A')$, as desired.

Suppose for the sake of a contradiction that $\widehat{\mu}(A) = +\infty$ for some measurable set $A \subseteq X$. Hence there are measurable sets $B \subseteq A$ such that $\|\mu(B)\|$ is as large as we want. Because $\mu(A)$ is equal to the sum of $\mu(B)$ and $\mu(A \setminus B)$, it follows that $\|\mu(B)\|$ and $\|\mu(A \setminus B)\|$ can both be as large as we want at the same time. Using the finite subadditivity of $\widehat{\mu}$ discussed in the previous paragraph, we get that $\widehat{\mu}(B) = +\infty$ or $\widehat{\mu}(A \setminus B) = +\infty$. By taking $C_1 = B$ or $A \setminus B$, as appropriate, we get a measurable subset of A such that $\widehat{\mu}(C_1) = +\infty$ and $\|\mu(C_1)\|$ is as large as we like. Repeating the process, we get a decreasing sequence of measurable sets C_1, C_2, \dots such that $\widehat{\mu}(C_l) = +\infty$ for each $l \geq 1$ and $\|\mu(C_l)\| \rightarrow \infty$ as $l \rightarrow \infty$.

This contradicts the fact that $\{\mu(C_l)\}_{l=1}^\infty$ converges in V to $\mu(\bigcap_{l=1}^\infty C_l)$, by the continuity condition that follows from countable additivity.

Let E be a nonempty set, and let $f(x)$ be a V -valued function on E such that $\sum_{x \in E} f(x)$ converges in the generalized sense. In particular, $\sum_{x \in E} f(x)$ satisfies the generalized Cauchy condition, and so for each $\epsilon > 0$ there is a finite set $B_\epsilon \subseteq E$ such that

$$(37.17) \quad \left\| \sum_{x \in C} f(x) \right\| < \epsilon$$

for every nonempty finite set $C \subseteq X \setminus B_\epsilon$. It follows that $\sum_{x \in A} f(x)$ satisfies the generalized Cauchy condition for every nonempty set $A \subseteq X$, since we can use $A \cap B_\epsilon$ in place of B_ϵ for the sum over A . Hence $\sum_{x \in A} f(x)$ converges in the generalized sense for every nonempty set $A \subseteq E$, because V is complete. Put

$$(37.18) \quad \mu(A) = \sum_{x \in A} f(x)$$

for each $A \subseteq E$, which is interpreted as being 0 when $A = \emptyset$. It is easy to see that this is a finitely-additive V -valued measure on the algebra of all subsets of E . Note that

$$(37.19) \quad \|\mu(C)\| \leq \epsilon$$

for every $C \subseteq X \setminus B_\epsilon$, since we can reduce to the previous case by approximating C by finite sets. Using this, one can check that μ is countably-additive. If $\|f(x)\|$ is a summable function on E , then μ satisfies the additional absolute convergence condition mentioned at the beginning of the section.

38 The Radon–Nikodym theorem

Let (X, \mathcal{A}) be a measurable space, and let μ, ν be finite nonnegative measures on (X, \mathcal{A}) such that

$$(38.1) \quad \mu(A) \leq C \nu(A)$$

for some $C \geq 0$ and every measurable set $A \subseteq X$. A special case of the Radon–Nikodym theorem states that there is a bounded nonnegative measurable function h on X such that

$$(38.2) \quad \mu(A) = \int_A h \, d\nu$$

for every measurable set $A \subseteq X$. Von Neumann's trick for showing this is to observe first that

$$(38.3) \quad \lambda(f) = \int_X f \, d\mu$$

is a bounded linear functional on $L^2(\nu)$. More precisely,

$$(38.4) \quad |\lambda(f)| \leq \int_X |f| \, d\mu \leq C \int_X |f| \, d\nu \leq C \nu(X)^{1/2} \left(\int_X |f|^2 \, d\nu \right)^{1/2},$$

using our hypothesis on μ and ν in the second step, and the Cauchy–Schwarz inequality in the third step. Because $L^2(X, \nu)$ is a Hilbert space, the Riesz representation theorem implies that there is an $h \in L^2(X, \nu)$ such that

$$(38.5) \quad \lambda(f) = \int_X f h \, d\nu$$

for every $f \in L^2(X, \nu)$. Hence

$$(38.6) \quad \mu(A) = \lambda(1_A) = \int_A h \, d\nu$$

for every measurable set $A \subseteq X$. It follows that

$$(38.7) \quad h(x) \leq C$$

almost everywhere on X with respect to ν under these conditions.

Instead of (38.1), suppose now that $\mu(A) = 0$ for every measurable set $A \subseteq X$ such that $\nu(A) = 0$. In this case, μ is said to be *absolutely continuous* with respect to ν , denoted $\mu \ll \nu$. The Radon–Nikodym theorem states that there is then a nonnegative measurable function h on X such that (38.2) holds for every measurable set $A \subseteq X$. More precisely, h is also integrable with respect to ν , because

$$(38.8) \quad \int_X h \, d\nu = \mu(X) < \infty.$$

To see this, we apply the previous version to μ and $\nu_1 = \mu + \nu$, since

$$(38.9) \quad \mu(A) \leq \mu(A) + \nu(A) = \nu_1(A)$$

for every measurable set $A \subseteq X$ trivially. This leads to a real-valued measurable function h_1 on X such that $0 \leq h_1 \leq 1$ and

$$(38.10) \quad \mu(A) = \int_A h_1 \, d\nu_1$$

for every measurable set A . If

$$(38.11) \quad B = \{x \in X : h_1(x) = 1\},$$

then B is measurable, and

$$(38.12) \quad \mu(B) = \nu_1(B) = \mu(B) + \nu(B),$$

which implies that $\nu(B) = 0$, and hence $\mu(B) = 0$. Thus $h_1 < 1$ ν -almost everywhere, and one may as well take h_1 so that $0 \leq h_1 < 1$ everywhere on X . If $A \subseteq X$ is measurable, then

$$(38.13) \quad \mu(A) = \int_A h_1 \, d\mu + \int_A h_1 \, d\nu$$

implies that

$$(38.14) \quad \int_A (1 - h_1) d\mu = \int_A h_1 d\nu,$$

and one can show that (38.2) holds with $h = h_1/(1 - h_1)$. More precisely,

$$(38.15) \quad \int_X g(1 - h_1) d\mu = \int_X g h_1 d\nu$$

for every bounded measurable function g on X , because of (38.14). If $h_1 \leq 1 - \delta$ on A for some $\delta > 0$, then one can take $g = 1/(1 - h_1)$ on A , $g = 0$ on $X \setminus A$, to get (38.2). One can then use countable additivity to get (38.2) for arbitrary measurable sets A .

If μ is a real or complex measure on (X, \mathcal{A}) , and not necessarily positive, then μ is still said to be absolutely continuous with respect to ν when $\mu(A) = 0$ for every measurable set $A \subseteq X$ such that $\nu(A) = 0$. This is equivalent to the condition that the total variation measure $|\mu|$ be absolutely continuous with respect to ν , which implies that μ can be expressed as a linear combination of finite nonnegative measures on X that are absolutely continuous with respect to ν . It follows from the previous case that there is a real or complex-valued integrable function h on X with respect to ν for which (38.2) holds. One can also allow ν to be σ -finite, by decomposing the domain into a countable union of pairwise-disjoint measurable sets of finite ν -measure. It is better to do this first when μ is nonnegative, to get the integrability of the density h , and then deal with real or complex measures μ .

Note that h is determined ν -almost everywhere by μ . More precisely, if h is a real or complex-valued integrable function on X with respect to ν such that

$$(38.16) \quad \int_A h d\nu = 0$$

for every measurable set $A \subseteq X$, then $h(x) = 0$ for almost every $x \in X$ with respect to ν . In the real case, one can simply take A to be the set where $h(x) > 0$ or $h(x) < 0$. The complex case follows from the real case, by considering the real and imaginary parts of h separately. If h' , h'' are integrable functions on X with respect to ν such that

$$(38.17) \quad \int_A h' d\nu = \int_A h'' d\nu$$

for every measurable set $A \subseteq X$, then it follows that $h = h' - h''$ is equal to 0 almost everywhere on X with respect to ν .

Of course, any real or complex measure μ on X is absolutely continuous with respect to the corresponding total variation measure $|\mu|$. The Radon–Nikodym theorem implies that there is an integrable function h on X with respect to $|\mu|$ such that

$$(38.18) \quad \mu(A) = \int_A h d|\mu|$$

for every measurable set $A \subseteq X$. Clearly

$$(38.19) \quad |\mu(A)| \leq \int_A |h| d|\mu|$$

for every measurable set $A \subseteq X$, which implies that

$$(38.20) \quad |\mu|(A) \leq \int_A |h| d|\mu|,$$

since the right side is a nonnegative measure on X . It follows that $|h(x)| \geq 1$ for almost every $x \in X$ with respect to $|\mu|$, and we would like to check that $|h(x)| = 1$ almost everywhere on X .

If μ is real and $A_1 = \{x \in X : h(x) > 1\}$ has positive $|\mu|$ -measure, then

$$(38.21) \quad \mu(A_1) = \int_{A_1} h d|\mu| > |\mu|(A_1) \geq \mu(A_1),$$

a contradiction. Thus $|\mu|(A_1) = 0$, and $|\mu|(\{x \in X : h(x) < -1\}) = 0$ for similar reasons. In the complex case, put $A_\alpha = \{x \in X : \operatorname{Re}(\alpha h(x)) > 1\}$ for each $\alpha \in \mathbf{C}$ with $|\alpha| = 1$. If $|\mu|(A_\alpha) > 0$ for some α , then

$$(38.22) \quad |\mu(A_\alpha)| \geq \operatorname{Re}(\alpha \mu(A_\alpha)) = \int_{A_\alpha} \operatorname{Re}(\alpha h) d|\mu| > |\mu|(A_\alpha) \geq |\mu(A_\alpha)|,$$

which is a contradiction again. This shows that $|\mu|(A_\alpha) = 0$ for every complex number α with $|\alpha| = 1$. Let $\{\alpha_j\}_{j=1}^\infty$ be a sequence of complex numbers with $|\alpha_j| = 1$ for each j which is dense in the unit circle in \mathbf{C} , such as an enumeration of the points on the circle that correspond to angles that are rational multiples of 2π . If $x \in X$ and $|h(x)| > 1$, then $x \in A_{\alpha_j}$ when α_j is sufficiently close to $\overline{h(x)}/|h(x)|$. Equivalently,

$$(38.23) \quad \{x \in X : |h(x)| > 1\} = \bigcup_{j=1}^\infty A_{\alpha_j},$$

and so $|\mu|(\{x \in X : |h(x)| > 1\}) = 0$, as desired. In particular, $h(x) = \pm 1$ almost everywhere on X with respect to $|\mu|$ in the real case, which implies the Hahn decomposition, as in Section 36.

39 The Lebesgue decomposition

Let (X, \mathcal{A}) be a measurable space, and let μ and ν be positive finite measures on X . If $\nu_1 = \mu + \nu$, then $\mu \leq \nu_1$, and there is a real-valued measurable function h_1 on X that satisfies $0 \leq h_1 \leq 1$ and (38.10), as before. Let B be as in (38.11), so that B is measurable and satisfies (38.12), which implies that $\nu(B) = 0$. However, without the additional hypothesis of absolute continuity of μ with

respect to ν , we do not necessarily have that $\mu(B) = 0$. Instead, let μ' , μ'' be the measures defined by

$$(39.1) \quad \mu'(A) = \mu(A \cap B), \quad \mu''(A) = \mu(A \cap (X \setminus B)).$$

By construction, μ' and ν are mutually singular, in the sense that $\nu(B) = 0$ and $\mu'(X \setminus B) = 0$. We still have (38.13), (38.14), and (38.15), which imply that

$$(39.2) \quad \mu''(A) = \int_{A \cap (X \setminus B)} \frac{h_1}{1 - h_1} d\nu$$

for every measurable set $A \subseteq X$. In particular, μ'' is absolutely continuous with respect to ν . Of course, $\mu = \mu' + \mu''$, which is known as the *Lebesgue decomposition* of μ . If μ is a real or complex measure on X , then an analogous decomposition can be obtained by applying this argument to $|\mu|$ in place of μ .

40 The Riesz representation theorem

Let (X, \mathcal{A}, μ) be a measure space, and let $1 \leq p, q \leq \infty$ be conjugate exponents, so that $1/p + 1/q = 1$. If $f \in L^p(X)$ and $g \in L^q(X)$, then the integral version of Hölder's inequality implies that $f g \in L^1(X)$, and that

$$(40.1) \quad \|f g\|_1 \leq \|f\|_p \|g\|_q.$$

The proof is basically the same as for sums, as in Section 20. It follows that

$$(40.2) \quad \lambda_g(f) = \int_X f g d\mu$$

defines a bounded linear functional on $L^p(X)$ when $g \in L^q(X)$, with dual norm less than or equal to $\|g\|_q$. If $p = \infty$, then it is easy to see that the dual norm of λ_g is equal to $\|g\|_1$, by choosing $f \in L^\infty(X)$ such that $\|f\|_\infty = 1$ and $f g = |g|$. Similarly, if $1 < p < \infty$, then the dual norm of λ_g on $L^p(X)$ is equal to $\|g\|_q$, because there is an $f \in L^p(X)$ such that $f g = |f|^p = |g|^q$. The dual norm of λ_g on $L^1(X)$ is also equal to $\|g\|_\infty$, under an additional hypothesis. More precisely, we should ask that for each measurable set $A \subseteq X$ with $\mu(A) > 0$ there is a measurable set $B \subseteq A$ such that $0 < \mu(B) < \infty$. This condition holds when μ is σ -finite on X , and for counting measure on any set X . If $0 \leq t < \|g\|_\infty$, then we can apply this to $A_t = \{x \in X : |g(x)| \geq t\}$ to get a measurable set $B_t \subseteq A_t$ with $0 < \mu(B_t) < \infty$. Put $f_t(x) = g(x)/|g(x)|$ for every $x \in B_t$ when g is real-valued, $f_t(x) = \overline{g(x)}/|g(x)|$ for every $x \in B_t$ when g is complex-valued, and $f_t(x) = 0$ for every $x \in X \setminus B_t$ in both cases. It is easy to see that $f_t \in L^1(X)$, $\|f_t\|_1 = \mu(B_t)$, and $\lambda_g(f_t) \geq t \mu(B_t)$, which implies that the dual norm of λ_g on $L^1(X)$ is greater than or equal to t . It follows that the dual norm of λ_g on $L^1(X)$ is greater than or equal to $\|g\|_\infty$, since this holds for every nonnegative real number t such that $t < \|g\|_\infty$. Hence the dual norm of λ_g on $L^1(X)$ is equal to $\|g\|_\infty$, since we already know that it is less than or equal to $\|g\|_\infty$.

Conversely, every bounded linear functional on $L^p(X)$ can be realized in this way when $1 < p < \infty$, and also when $p = 1$ and X has σ -finite μ -measure. To see this, let us begin with the case where $\mu(X) < \infty$. Let λ be a bounded linear functional on $L^p(X)$, $1 \leq p < \infty$, and put

$$(40.3) \quad \nu(A) = \lambda(\mathbf{1}_A)$$

for every measurable set $A \subseteq X$. Here $\mathbf{1}_A$ denotes the indicator function on X associated to A , equal to 1 on A and to 0 on $X \setminus A$. If A_1, A_2, \dots is a sequence of pairwise-disjoint measurable subsets of X , then $\sum_{j=1}^{\infty} \mathbf{1}_{A_j}$ converges in $L^p(X)$ to the indicator function associated to $\bigcup_{j=1}^{\infty} A_j$ when $p < \infty$, and hence

$$(40.4) \quad \nu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \nu(A_j).$$

Thus ν is a real or complex measure on X , as appropriate. This measure is also absolutely continuous with respect to μ , since $\mathbf{1}_A = 0$ in $L^p(X)$ when $\mu(A) = 0$. The Radon–Nikodym theorem implies that there is a $g \in L^1(X)$ such that

$$(40.5) \quad \nu(A) = \int_A g \, d\mu$$

for every measurable set $A \subseteq X$. By linearity, it follows that

$$(40.6) \quad \lambda(f) = \int_X f g \, d\mu$$

for every measurable simple function f on X . This also holds when f is a bounded measurable function on X , by approximating f by simple functions. If $p = 1$, then one can use this to show that $g \in L^\infty(X)$, with L^∞ norm less than or equal to the dual norm of λ on $L^1(X)$, in the same way as in the previous paragraph. If $p > 1$, then one can first show that the L^q norm of the restriction of g to any set on which it is bounded is less than or equal to the dual norm of λ on $L^p(X)$, by the same type of argument as in the previous paragraph. This implies that $g \in L^q(X)$, with L^q norm less than or equal to the dual norm of λ on $L^p(X)$. In both cases, one can then use the boundedness of λ on $L^p(X)$ and the fact that $g \in L^q(X)$ to show that (40.6) holds for every $f \in L^p(X)$, because simple functions are dense in $L^p(X)$.

Suppose now that X has σ -finite μ -measure, so that there is a sequence of measurable subsets E_1, E_2, \dots of X such that $\mu(E_l) < \infty$ for each $l \geq 1$ and $\bigcup_{l=1}^{\infty} E_l = X$. We may also suppose that $E_k \cap E_l = \emptyset$ when $k \neq l$, by replacing E_l with $E_l \setminus (E_1 \cup \dots \cup E_{l-1})$ when $l > 1$. If λ is a bounded linear functional on $L^p(X)$, then the restriction of λ to $f \in L^p(X)$ such that $f = 0$ on $X \setminus E_l$ defines a bounded linear functional on $L^p(E_l)$ for each l . By the previous argument, for each positive integer l , there is a $g_l \in L^q(E_l)$ such that

$$(40.7) \quad \lambda(f) = \int_{E_l} f g_l \, d\mu$$

for every $f \in L^p(E_l)$. Let g be the function on X defined by $g = g_l$ on E_l for each l . Thus the restriction of g to $\bigcup_{j=1}^n E_l$ is in L^q for each n , and $\lambda(f)$ is equal to the integral of f times g when $f \in L^p(X)$ and $f = 0$ on $X \setminus \left(\bigcup_{l=1}^n E_l\right)$. In particular, the L^q norm of the restriction of g to $\bigcup_{l=1}^n E_l$ is less than or equal to the dual norm of the restriction of λ to $L^p\left(\bigcup_{l=1}^n E_l\right)$ for each n , which is bounded by the dual norm of λ on $L^p(X)$. This implies that $g \in L^q(X)$, with L^q norm less than or equal to the dual norm of λ on $L^p(X)$. Every $f \in L^p(X)$ can be approximated in the L^p norm by functions that are equal to 0 on $X \setminus \left(\bigcup_{l=1}^n E_l\right)$ for some n , because $q < \infty$, and so $\lambda(f)$ is given by the integral of f times g for every $f \in L^p(X)$.

If $1 < p < \infty$, then we can drop the hypothesis that X be σ -finite. To see this, let a bounded linear functional λ on $L^p(X)$ be given. We may as well suppose that $\lambda \neq 0$, since otherwise there is nothing to do. In particular, $L^p(X) \neq \{0\}$, which is to say that there are measurable subsets of X with positive finite measure. If $Y \subseteq X$ is measurable and σ -finite, then there is a $g_Y \in L^q(Y)$ such that

$$(40.8) \quad \lambda(f) = \int_Y f g_Y d\mu$$

for every $f \in L^p(X)$ with $f = 0$ on $X \setminus Y$, by the previous argument. Moreover, the L^q norm of g_Y is equal to the dual norm of the restriction of λ to $L^p(Y)$, which is less than or equal to the dual norm of λ on $L^p(X)$. Let f_1, f_2, \dots be a sequence of elements of $L^p(X)$ such that $\|f_j\|_p = 1$ for each j and $\{|\lambda(f_j)|\}_{j=1}^\infty$ converges to the dual norm of λ on $L^p(X)$. Observe that

$$(40.9) \quad Y_0 = \bigcup_{j=1}^\infty \{x \in X : f_j(x) \neq 0\}$$

is a measurable set with σ -finite measure, because the set where $f_j \neq 0$ has this property for each j . Hence there is a $g_{Y_0} \in L^q(Y_0)$ with the properties mentioned earlier. By construction, the dual norm of λ on $L^p(X)$ is equal to the dual norm of the restriction of λ to $L^p(Y_0)$, which is equal to the L^q norm of g_{Y_0} . If $Y \subseteq X$ is measurable and σ -finite, and if $Y_0 \subseteq Y$, then $g_Y = g_{Y_0}$ almost everywhere on Y_0 , by uniqueness of the representation. However, the L^q norm of g_Y is less than or equal to the dual norm of λ on $L^p(X)$, which is equal to the L^q norm of g_{Y_0} . This implies that $g_Y = 0$ almost everywhere on $Y \setminus Y_0$, since $q < \infty$. Let g be the function on X equal to g_{Y_0} on Y_0 and to 0 on $X \setminus Y_0$. If $f \in L^p(X)$, then the previous argument can be applied to

$$(40.10) \quad Y = Y_0 \cup \{x \in X : f(x) \neq 0\},$$

to get that $\lambda(f)$ is equal to the integral of f times g , as desired.

41 Lengths of paths

Let $(M, d(x, y))$ be a metric space, and let f be a function on a closed interval $[a, b]$ in the real line with values in M . If $\mathcal{P} = \{t_j\}_{j=0}^n$ is a partition of $[a, b]$, in the sense that

$$(41.1) \quad a = t_0 < t_1 < \cdots < t_n = b,$$

then we put

$$(41.2) \quad \Lambda_a^b(\mathcal{P}) = \sum_{j=1}^n d(f(t_j), f(t_{j-1})).$$

Note that

$$(41.3) \quad d(f(a), f(b)) \leq \Lambda_a^b(\mathcal{P}),$$

because of the triangle inequality. Similarly,

$$(41.4) \quad \Lambda_a^b(\mathcal{P}) \leq \Lambda_a^b(\mathcal{P}')$$

when \mathcal{P}' is another partition of $[a, b]$ that is a refinement of \mathcal{P} , which means that \mathcal{P}' includes the points in \mathcal{P} . The length Λ_a^b of the path $f(t)$, $a \leq t \leq b$, is defined to be the supremum of $\Lambda_a^b(\mathcal{P})$ over all partitions \mathcal{P} of $[a, b]$, which may be infinite.

Suppose that $a \leq r \leq b$, and that $\mathcal{P}_1, \mathcal{P}_2$ are partitions of $[a, r], [r, b]$, respectively. We can combine $\mathcal{P}_1, \mathcal{P}_2$ to get a partition \mathcal{P} of $[a, b]$ that satisfies

$$(41.5) \quad \Lambda_a^r(\mathcal{P}_1) + \Lambda_r^b(\mathcal{P}_2) = \Lambda_a^b(\mathcal{P}).$$

Thus

$$(41.6) \quad \Lambda_a^r(\mathcal{P}_1) + \Lambda_r^b(\mathcal{P}_2) \leq \Lambda_a^b,$$

which implies that

$$(41.7) \quad \Lambda_a^r + \Lambda_r^b \leq \Lambda_a^b,$$

by taking the supremum over all partitions $\mathcal{P}_1, \mathcal{P}_2$ of $[a, r], [r, b]$. In the other direction, if \mathcal{P} is any partition of $[a, b]$, then \mathcal{P} may or may not include r , but we can add r to \mathcal{P} if necessary to get a refinement \mathcal{P}' of \mathcal{P} that does contain r . This permits \mathcal{P}' to be expressed as the combination of partitions $\mathcal{P}_1, \mathcal{P}_2$ of $[a, r], [r, b]$, respectively, so that

$$(41.8) \quad \Lambda_a^b(\mathcal{P}) \leq \Lambda_a^b(\mathcal{P}') = \Lambda_a^r(\mathcal{P}_1) + \Lambda_r^b(\mathcal{P}_2).$$

Hence

$$(41.9) \quad \Lambda_a^b(\mathcal{P}) \leq \Lambda_a^r + \Lambda_r^b$$

for every partition \mathcal{P} of $[a, b]$, and therefore

$$(41.10) \quad \Lambda_a^b \leq \Lambda_a^r + \Lambda_r^b.$$

Combining this with (41.7), we get that

$$(41.11) \quad \Lambda_a^b = \Lambda_a^r + \Lambda_r^b.$$

In particular,

$$(41.12) \quad \Lambda_a^r \leq \Lambda_a^b$$

when $a \leq r \leq b$, which can be seen more directly by extending any partition of $[a, r]$ to a partition of $[a, b]$.

The diameter of a nonempty set $E \subseteq M$ is defined by

$$(41.13) \quad \text{diam } E = \sup\{d(x, y) : x, y \in E\},$$

which is finite exactly when E is bounded. If $a \leq r \leq t \leq b$ and \mathcal{P} is a partition of $[a, b]$ consisting of these points, then

$$(41.14) \quad d(f(r), f(t)) \leq \Lambda_a^b(\mathcal{P}) \leq \Lambda_a^b.$$

It follows that

$$(41.15) \quad \text{diam } f([a, b]) \leq \Lambda_a^b.$$

Note that $\Lambda_a^b = 0$ if and only if f is constant.

Consider the special case where $M = \mathbf{R}$ and $f : [a, b] \rightarrow \mathbf{R}$ is monotone increasing. If $\mathcal{P} = \{t_j\}_{j=0}^n$ is any partition of $[a, b]$, then

$$(41.16) \quad \Lambda_a^b(\mathcal{P}) = \sum_{j=1}^n (f(t_j) - f(t_{j-1})) = f(b) - f(a).$$

This implies that

$$(41.17) \quad \Lambda_a^b = f(b) - f(a).$$

42 Lipschitz mappings

Let $(M_1, d_1(x, y))$ and $(M_2, d_2(u, v))$ be metric spaces. A mapping $f : M_1 \rightarrow M_2$ is said to be *Lipschitz* if there is a constant $k \geq 0$ such that

$$(42.1) \quad d_2(f(x), f(y)) \leq k d_1(x, y)$$

for every $x, y \in M_1$. Thus Lipschitz mappings are automatically uniformly continuous, and f is Lipschitz with $k = 0$ if and only if f is constant.

If M_2 is the real line with the standard metric, then $f : M_1 \rightarrow \mathbf{R}$ is Lipschitz with constant k if and only if

$$(42.2) \quad f(x) \leq f(y) + k d_1(x, y)$$

for every $x, y \in M_1$. More precisely, (42.1) implies (42.2) directly, and to get the converse, one can apply the latter both to x, y and with the roles of x, y exchanged. In particular,

$$(42.3) \quad f_p(x) = d_1(x, p)$$

is Lipschitz with constant 1 on M_1 for every $p \in M_1$, by the triangle inequality. For example, $f(x) = |x|$ is Lipschitz with constant 1 on the real line.

Suppose now that f is a Lipschitz mapping with constant k from a closed interval $[a, b]$ in the real line with the standard metric into a metric space $(M_2, d_2(u, v))$. If $\mathcal{P} = \{t_j\}_{j=0}^n$ is a partition of $[a, b]$, then

$$(42.4) \quad \Lambda_a^b(\mathcal{P}) = \sum_{j=1}^n d_2(f(t_j), f(t_{j-1})) \leq \sum_{j=1}^n k(t_j - t_{j-1}) = k(b - a).$$

Thus f has length $\Lambda_a^b \leq k(b - a)$.

If M_1 , M_2 , and M_3 are metric spaces, and $f_1 : M_1 \rightarrow M_2$, $f_2 : M_2 \rightarrow M_3$ are Lipschitz mappings with constants k_1 , k_2 , respectively, then their composition $f_2 \circ f_1$ is a Lipschitz mapping from M_1 into M_3 with constant $k_1 k_2$. Similarly, if $f_1 : [a, b] \rightarrow M_2$ has length Λ_a^b and $f_2 : M_2 \rightarrow M_3$ is Lipschitz with constant k_2 , then $f_2 \circ f_1 : [a, b] \rightarrow M_3$ has length $\leq k_2 \Lambda_a^b$.

43 Bounded variation

A real-valued function f on a closed interval $[a, b]$ in the real line is said to have *bounded variation* if it has finite length as a mapping into \mathbf{R} with the standard metric. In this case, the length of f is also known as its *total variation*. We can also consider the positive and negative variations of f separately, as follows.

For each real number x , put $x_+ = x$ when $x \geq 0$, $x_+ = 0$ when $x \leq 0$, $x_- = -x$ when $x \leq 0$, and $x_- = 0$ when $x \geq 0$. Thus

$$(43.1) \quad x_+ + x_- = |x|, \quad x_+ - x_- = x$$

and

$$(43.2) \quad (x + y)_+ \leq x_+ + y_+, \quad (x + y)_- \leq x_- + y_-$$

for every $x, y \in \mathbf{R}$. If $\mathcal{P} = \{t_j\}_{j=0}^n$ is a partition of $[a, b]$, then put

$$(43.3) \quad P_a^b(\mathcal{P}) = \sum_{j=1}^n (f(t_j) - f(t_{j-1}))_+$$

and

$$(43.4) \quad N_a^b(\mathcal{P}) = \sum_{j=1}^n (f(t_j) - f(t_{j-1}))_-.$$

Note that

$$(43.5) \quad P_a^b(\mathcal{P}) + N_a^b(\mathcal{P}) = \Lambda_a^b(\mathcal{P})$$

and

$$(43.6) \quad P_a^b(\mathcal{P}) - N_a^b(\mathcal{P}) = f(b) - f(a),$$

by (43.1). If \mathcal{P}' is another partition of $[a, b]$ which is a refinement of \mathcal{P} , then it is easy to see that

$$(43.7) \quad P_a^b(\mathcal{P}) \leq P_a^b(\mathcal{P}'), \quad N_a^b(\mathcal{P}) \leq N_a^b(\mathcal{P}'),$$

using (43.2).

Let P_a^b, N_a^b be the suprema of $P_a^b(\mathcal{P}), N_a^b(\mathcal{P})$ over all partitions \mathcal{P} of $[a, b]$, respectively. Clearly

$$(43.8) \quad \Lambda_a^b \leq P_a^b + N_a^b,$$

by (43.5). To get the opposite inequality

$$(43.9) \quad P_a^b + N_a^b \leq \Lambda_a^b,$$

one should be a bit more careful, because the partitions \mathcal{P} of $[a, b]$ for which $P_a^b(\mathcal{P})$ approaches P_a^b may not be the same as the partitions for which $N_a^b(\mathcal{P})$ approaches N_a^b . However, using common refinements of such partitions, one can get partitions \mathcal{P} such that $P_a^b(\mathcal{P}), N_a^b(\mathcal{P})$ approach P_a^b, N_a^b at the same time. This implies (43.9), from which it follows that

$$(43.10) \quad P_a^b + N_a^b = \Lambda_a^b.$$

Observe also that

$$(43.11) \quad P_a^r + P_r^b = P_a^b, \quad N_a^r + N_r^b = N_a^b$$

for each $r, a \leq r \leq b$. This uses the same arguments as for Λ_a^b , in Section 41.

Suppose now that f has bounded variation, so that $\Lambda_a^b < \infty$, and hence $P_a^b, N_a^b < \infty$. Using (43.6), one can check that

$$(43.12) \quad P_a^b - N_a^b = f(b) - f(a).$$

More precisely, one should be careful to use partitions \mathcal{P} of $[a, b]$ such that P_a^b, N_a^b are simultaneously approximated by $P_a^b(\mathcal{P}), N_a^b(\mathcal{P})$, respectively, as in the previous paragraph. Similarly,

$$(43.13) \quad P_a^r - N_a^r = f(r) - f(a)$$

for each $r \in [a, b]$, since the restriction of f to $[a, r]$ also has bounded variation. Of course, P_a^r and N_a^r are monotone increasing on $[a, b]$.

44 Functions and measures

Let $\alpha(x)$ be a monotone increasing real-valued function on the real line. As usual, the one-sided limits $\alpha(x+) = \lim_{y \rightarrow x+} \alpha(y)$, $\alpha(x-) = \lim_{z \rightarrow x-} \alpha(z)$ exist for every $x \in \mathbf{R}$, and are given by

$$(44.1) \quad \alpha(x+) = \sup\{\alpha(y) : y \in \mathbf{R}, y < x\},$$

$$(44.2) \quad \alpha(x-) = \inf\{\alpha(z) : z \in \mathbf{R}, x < z\}.$$

Thus

$$(44.3) \quad \alpha(x-) \leq \alpha(x) \leq \alpha(x+)$$

for every $x \in \mathbf{R}$, and $\alpha(x+) = \alpha(x-)$ exactly when α is continuous at x . Moreover,

$$(44.4) \quad \alpha(x+) \leq \alpha(y-)$$

for every $x, y \in \mathbf{R}$ with $x < y$. Remember that the set of $x \in \mathbf{R}$ at which α is not continuous has only finitely or countably many elements.

It is well known that there is a unique positive Borel measure μ_α on \mathbf{R} that satisfies

$$(44.5) \quad \mu_\alpha((a, b)) = \alpha(b-) - \alpha(a+), \quad \mu_\alpha([a, b]) = \alpha(b+) - \alpha(a-)$$

for every $a, b \in \mathbf{R}$ with $a < b$. The expression for closed intervals also makes sense when $a = b$, in which case it reduces to

$$(44.6) \quad \alpha(\{a\}) = \alpha(a+) - \alpha(a-).$$

Of course, this is equal to 0 when α is continuous at a . Alternatively, if f is a continuous real-valued function on the real line with compact support, then one can define the Riemann–Stieltjes integral

$$(44.7) \quad \int_{-\infty}^{\infty} f(x) d\alpha(x).$$

This is a nonnegative linear functional on the space of continuous functions with compact support on \mathbf{R} , and the Riesz representation theorem leads to a positive Borel measure that is the same as μ_α . As another approach, if α is a strictly increasing continuous function on \mathbf{R} , then one can get μ_α from Lebesgue measure using a change of variables. If α is monotone increasing and continuous, but perhaps not strictly increasing, then

$$(44.8) \quad \beta(x) = \alpha(x) + x$$

is continuous and strictly increasing, the previous argument can be used to get μ_β , and one can get μ_α by subtracting Lebesgue measure from μ_β . If α is not continuous, then one can account for the discontinuities directly with sums of multiples of Dirac masses.

Let us say that a real-valued function α on \mathbf{R} has bounded variation if it has bounded variation on every closed interval $[a, b]$, and if the total variation Λ_a^b of α on $[a, b]$ is uniformly bounded. This implies that α is bounded on \mathbf{R} , since

$$(44.9) \quad |\alpha(a) - \alpha(b)| \leq \Lambda_a^b$$

for every $a, b \in \mathbf{R}$ with $a \leq b$. It is easy to see that bounded monotone functions on \mathbf{R} have bounded variation. Conversely, one can check that a function with bounded variation on \mathbf{R} can be expressed as a difference of monotone increasing functions that are bounded. Complex-valued functions of bounded variation on \mathbf{R} can be defined analogously, and represented as linear combinations of bounded monotone real-valued functions.

If α is a real or complex-valued function of bounded variation on \mathbf{R} , then there is a real or complex measure Borel measure μ_α on \mathbf{R} associated to α as before. More precisely, if α is given as a linear combination of bounded monotone increasing real-valued functions, then μ_α is the same as the corresponding linear

combination of positive finite measures. In this case, the Riemann-Stieltjes integral (44.7) defines a bounded linear functional on the space of continuous functions on \mathbf{R} with compact support with respect to the supremum norm, which leads to a real or complex Borel measure on \mathbf{R} , as appropriate.

45 Continuity conditions

Let $(M, d(x, y))$ be a complete metric space, and let $f : [a, b] \rightarrow M$ be a path of finite length Λ_a^b . If $\{t_j\}_{j=1}^\infty$ is a monotone sequence of elements of $[a, b]$, then it is easy to see that

$$(45.1) \quad \sum_{j=1}^n d(f(t_j), f(t_{j+1})) \leq \Lambda_a^b$$

for every positive integer n . This implies that $\sum_{j=1}^\infty d(f(t_j), f(t_{j+1}))$ converges, and hence that $\{f(t_j)\}_{j=1}^\infty$ converges in M , as in Section 11. Using this, one can check that $f(r+) = \lim_{t \rightarrow r+} f(t)$ exists for every $r \in [a, b)$, and similarly that $f(r-) = \lim_{t \rightarrow r-} f(t)$ exists for every $r \in (a, b]$. More precisely, this also uses the observation that two strictly increasing or two strictly decreasing sequences with the same limit can be combined into a single monotone sequence, and hence that the corresponding sequences of values of f have the same limit in M .

Alternatively, let Λ_u^v be the length of the restriction of f to $[u, v]$ when $a \leq u \leq v \leq b$. Of course, Λ_a^r is monotone increasing in r , and hence

$$(45.2) \quad \lim_{t \rightarrow r-} \Lambda_a^t = \sup_{a \leq t < r} \Lambda_a^t$$

when $a < r \leq b$. Let $\epsilon > 0$ be given, and choose $u \in [a, r)$ so that

$$(45.3) \quad \Lambda_a^u > \sup_{a \leq t < r} \Lambda_a^t - \epsilon.$$

Because $\Lambda_a^t = \Lambda_a^u + \Lambda_u^t$ when $u \leq t < r$, we get that

$$(45.4) \quad \sup_{u \leq t < r} \Lambda_u^t < \epsilon.$$

One can also use this to deal with $f(r-)$, and similarly for $f(r+)$ when $a \leq r < b$.

If $a \leq r \leq t \leq b$, then

$$(45.5) \quad d(f(r), f(t)) \leq \Lambda_r^t,$$

as usual. It follows that f is continuous on the right at $r \in [a, b)$ when

$$(45.6) \quad \lim_{t \rightarrow r+} \Lambda_r^t = 0,$$

and that f is continuous from the left at $r \in (a, b]$ when

$$(45.7) \quad \lim_{t \rightarrow r-} \Lambda_t^r = 0.$$

Equivalently, continuity of Λ_a^r from the right or the left implies continuity of $f(r)$ from the right or the left at the same point, respectively. In particular, $f(r)$ is continuous at every point where Λ_a^r is continuous, which includes all but at most finitely or countably many elements of $[a, b]$, because Λ_a^r is monotone increasing in r .

Conversely, Λ_a^r is continuous from the right or left at any point where f is continuous from the right or left. To see this, let $r \in (a, b]$ and $\epsilon > 0$ be given, and let $\mathcal{P} = \{t_j\}_{j=0}^n$ be a partition of $[a, r]$ such that

$$(45.8) \quad \Lambda_a^r(\mathcal{P}) > \Lambda_a^r - \epsilon.$$

If $t_{n-1} < t < t_n = r$, then let \mathcal{P}_t be the partition of $[a, r]$ obtained by adding t between t_{n-1} and $t_n = r$ in \mathcal{P} . Thus \mathcal{P}_t is a refinement of \mathcal{P} , so that

$$(45.9) \quad \Lambda_a^r(\mathcal{P}_t) \geq \Lambda_a^r(\mathcal{P}).$$

We can also consider \mathcal{P}_t as the combination of a partition of $[0, t]$ with a single step from t to r , which implies that

$$(45.10) \quad \Lambda_a^r(\mathcal{P}_t) \leq \Lambda_a^t + d(f(t), f(r)).$$

Hence

$$(45.11) \quad \Lambda_a^t + d(f(t), f(r)) > \Lambda_a^r - \epsilon$$

when $t_{n-1} < t < r$. This shows that Λ_a^r is continuous from the left at r when $f(r)$ is continuous from the left at r , using also the fact that $\Lambda_a^t \leq \Lambda_a^r$ when $a \leq t \leq r$. The argument for continuity on the right is very similar.

46 Maximal functions

Let μ be a positive finite Borel measure on the real line. The *Hardy–Littlewood maximal function* associated to μ is defined by

$$(46.1) \quad \mu^*(x) = \sup_{x \in I} \frac{\mu(I)}{|I|},$$

where the supremum is taken over all open intervals (a, b) that contain x , and $|I| = b - a$ is the length of I . Put

$$(46.2) \quad E_t = \{x \in \mathbf{R} : \mu^*(x) > t\}$$

for each $t > 0$. Thus $x \in E_t$ if and only if there is an open interval I such that $x \in I$ and

$$(46.3) \quad \mu(I) > t |I|.$$

In this case, $I \subseteq E_t$, and it follows that E_t is an open set in \mathbf{R} .

Suppose that $K \subseteq E_t$ is compact. This implies that there are finitely many open intervals I_1, \dots, I_n in \mathbf{R} such that

$$(46.4) \quad K \subseteq \bigcup_{j=1}^n I_j$$

and

$$(46.5) \quad \mu(I_j) > t |I_j|$$

for each j . A basic property of the real line is that for any three intervals with a point in common, one of the intervals is contained in the union of the other two. This permits us to reduce the collection of intervals I_1, \dots, I_n in such a way that no element of \mathbf{R} is contained in more than two of these intervals.

It follows that

$$(46.6) \quad \sum_{j=1}^n |I_j| < t^{-1} \sum_{j=1}^n \mu(I_j) \leq 2 t^{-1} \mu\left(\bigcup_{j=1}^n I_j\right).$$

More precisely, if $\mathbf{1}_A$ is the indicator function on \mathbf{R} associated to $A \subseteq \mathbf{R}$, then

$$(46.7) \quad \sum_{j=1}^n \mu(I_j) = \int_{\mathbf{R}} \left(\sum_{j=1}^n \mathbf{1}_{I_j} \right) d\mu \leq \int_{\mathbf{R}} 2 \mathbf{1}_{\bigcup_{j=1}^n I_j} d\mu = 2 \mu\left(\bigcup_{j=1}^n I_j\right).$$

If $|K|$ denotes the Lebesgue measure of K , then we get that

$$(46.8) \quad |K| \leq 2 t^{-1} \mu(\mathbf{R}).$$

Hence

$$(46.9) \quad |E_t| \leq 2 t^{-1} \mu(\mathbf{R}),$$

because K is an arbitrary compact subset of E_t .

If f is an integrable function on \mathbf{R} , then we put

$$(46.10) \quad f^*(x) = \sup_{x \in I} \frac{1}{|I|} \int_I |f(y)| dy.$$

This is the same as the maximal function $\mu^*(x)$ associated to the measure

$$(46.11) \quad \mu(A) = \int_A |f(y)| dy.$$

Thus the estimate in the previous paragraph can be re-expressed in this case as

$$(46.12) \quad |\{x \in \mathbf{R} : f^*(x) > t\}| \leq 2 t^{-1} \int_{\mathbf{R}} |f(y)| dy$$

for each $t > 0$.

47 Lebesgue's theorem

Let f be a locally integrable function on the real line. A famous theorem of Lebesgue implies that

$$(47.1) \quad \lim_{r \rightarrow 0} \frac{1}{2r} \int_{x-r}^{x+r} |f(y) - f(x)| dy = 0$$

for almost every $x \in \mathbf{R}$. We may as well suppose that f is integrable on \mathbf{R} , since the problem is local.

Put

$$\begin{aligned} (47.2) \quad L(f)(x) &= \limsup_{r \rightarrow 0} \frac{1}{2r} \int_{x-r}^{x+r} |f(y) - f(x)| dy \\ &= \lim_{\epsilon \rightarrow 0} \sup_{0 < r < \epsilon} \frac{1}{2r} \int_{x-r}^{x+r} |f(y) - f(x)| dy. \end{aligned}$$

Observe that

$$(47.3) \quad L(f_1 + f_2)(x) \leq L(f_1)(x) + L(f_2)(x),$$

and that

$$(47.4) \quad L(g)(x) = 0$$

when g is continuous at x . It follows that

$$(47.5) \quad L(f) = L(f - g)$$

for every continuous function g .

We also have that

$$(47.6) \quad L(f)(x) \leq f^*(x) + |f(x)|,$$

where $f^*(x)$ is as in (46.10). This implies that

$$(47.7) \quad L(f)(x) \leq (f - g)^*(x) + |f(x) - g(x)|$$

for every continuous function g on \mathbf{R} . Hence

$$\begin{aligned} (47.8) \quad \{x \in \mathbf{R} : L(f)(x) > t\} \\ \subseteq \{x \in \mathbf{R} : (f - g)^*(x) > t/2\} \cup \{x \in \mathbf{R} : |f(x) - g(x)| > t/2\} \end{aligned}$$

for every $t > 0$.

As in the previous section,

$$(47.9) \quad |\{x \in \mathbf{R} : (f - g)^*(x) > t/2\}| \leq 2(t/2)^{-1} \|f - g\|_1 = 4t^{-1} \|f - g\|_1.$$

Similarly,

$$\begin{aligned} (47.10) \quad &|\{x \in \mathbf{R} : |f(x) - g(x)| > t/2\}| \\ &\leq (t/2)^{-1} \int_{\mathbf{R}} |f(y) - g(y)| dy = 2t^{-1} \|f - g\|_1 \end{aligned}$$

for every $t > 0$. Of course, we can choose g so that $\|f - g\|_1$ is arbitrarily small, because continuous functions are dense in $L^1(\mathbf{R})$. Using this, one can show that $L(f)(x) = 0$ almost everywhere, as desired.

48 Singular measures

Let μ be a positive finite Borel measure on the real line which is singular with respect to Lebesgue measure. This means that there is a Borel set $B \subseteq \mathbf{R}$ whose Lebesgue measure $|B|$ is 0 while $\mu(\mathbf{R} \setminus B) = 0$. Let us check that

$$(48.1) \quad \lim_{r \rightarrow 0} \frac{\mu((x-r, x+r))}{2r} = 0$$

for almost every $x \in \mathbf{R}$ with respect to Lebesgue measure. If B happens to be a closed set in \mathbf{R} , then this holds trivially for every $x \in \mathbf{R} \setminus B$. The idea is to use the maximal function to make an approximation by this type of situation.

Consider

$$(48.2) \quad \begin{aligned} L(\mu)(x) &= \limsup_{r \rightarrow 0} \frac{\mu((x-r, x+r))}{2r} \\ &= \lim_{\epsilon \rightarrow 0} \sup_{0 < r < \epsilon} \frac{\mu((x-r, x+r))}{2r}, \end{aligned}$$

in analogy with the previous section. Thus

$$(48.3) \quad L(\mu)(x) \leq \mu^*(x)$$

and

$$(48.4) \quad L(\mu_1 + \mu_2)(x) \leq L(\mu_1)(x) + L(\mu_2)(x)$$

for every pair of positive Borel measures μ_1, μ_2 on \mathbf{R} .

Let U be an open set in \mathbf{R} such that $|B| \subseteq U$, and let K be a compact set in \mathbf{R} such that $K \subseteq U$. Also let μ_1, μ_2 be the Borel measures on \mathbf{R} defined by

$$(48.5) \quad \mu_1(A) = \mu(A \cap K), \quad \mu_2(A) = \mu(A \cap (\mathbf{R} \setminus K)).$$

Thus $L(\mu_1)(x) = 0$ when $x \in \mathbf{R} \setminus K$, which implies that

$$(48.6) \quad L(\mu)(x) \leq L(\mu_2)(x) \leq \mu_2^*(x)$$

for every $x \in \mathbf{R} \setminus K$, and hence for every $x \in \mathbf{R} \setminus U$. The main point now is to choose $K \subseteq U$ so that

$$(48.7) \quad \mu(U \setminus K) = \mu(\mathbf{R} \setminus K) = \mu_2(\mathbf{R})$$

is arbitrarily small. This is easy to do, using the fact that open subsets of the real line are σ -compact. This implies that $L(\mu)(x) = 0$ for Lebesgue almost every $x \in \mathbf{R} \setminus U$, by the maximal function estimates in Section 46. More precisely,

$$(48.8) \quad \{x \in \mathbf{R} \setminus U : L(\mu)(x) > t\} \subseteq \{x \in \mathbf{R} : \mu_2^*(x) > t\}$$

for every $t > 0$, and the Lebesgue measure of the set on the right can be made arbitrarily small, by choosing K so that (48.7) is small. This implies that $L(\mu)(x) \leq t$ almost everywhere on $\mathbf{R} \setminus U$ with respect to Lebesgue measure for each $t > 0$, and hence that $L(\mu)(x) = 0$ almost everywhere on $\mathbf{R} \setminus U$, by taking $t = 1/n$, where n is a positive integer. It follows that $L(\mu)(x) = 0$ for Lebesgue almost every $x \in \mathbf{R}$, as desired, since we can also choose U to have arbitrarily small Lebesgue measure, because $|B| = 0$.

49 Differentiability almost everywhere

Let α be a bounded real-valued monotone increasing function on the real line, and let μ_α be the corresponding positive Borel measure on \mathbf{R} , as in Section 44. Using the Lebesgue decomposition and Radon–Nikodym theorem, we get an integrable function f with respect to Lebesgue measure and a Borel measure ν that is singular with respect to Lebesgue measure such that

$$(49.1) \quad \mu_\alpha(A) = \int_A f(y) dy + \nu(A).$$

We would like to show that $\alpha(x)$ is differentiable almost everywhere on \mathbf{R} with respect to Lebesgue measure, and more precisely that $\alpha'(x) = f(x)$ almost everywhere.

Thus we would like to show that

$$(49.2) \quad \lim_{h \rightarrow 0} \frac{\alpha(x+h) - \alpha(x)}{h} = f(x)$$

for almost every $x \in \mathbf{R}$. As a first approximation, we have that

$$(49.3) \quad \lim_{h \rightarrow 0} \frac{1}{h} \int_x^{x+h} f(y) dy = f(x)$$

for almost every $x \in \mathbf{R}$, by Lebesgue's theorem. More precisely, this integral is supposed to be oriented, as in calculus, so that the integral from x to $x+h$ is -1 times the integral from $x+h$ to x . This means that we are looking at the average of f over the interval $[x, x+h]$ when $h > 0$, and over $[x+h, x]$ when $h < 0$.

It remains to show that

$$(49.4) \quad \frac{\alpha(x+h) - \alpha(x)}{h} - \frac{1}{h} \int_x^{x+h} f(y) dy$$

converges to 0 as $h \rightarrow 0$ for almost every $x \in \mathbf{R}$. If α is continuous at x and $x+h$, then this difference is equal to $\nu([x, x+h])/h$ when $h > 0$, and similarly when $h < 0$. In any case, this difference is nonnegative, bounded by $\nu([x, x+h])/h$ when $h > 0$, and similarly for $h < 0$. Hence the difference converges to 0 almost everywhere, as in the previous section.

Of course, it is not important that α be bounded or defined on the whole line, since the problem is local. If α is a real or complex-valued function of bounded variation on \mathbf{R} , then α can be expressed as a linear combination of monotone functions, and is therefore differentiable almost everywhere too.

50 Maximal functions, 2

The maximal function of a positive Borel measure μ on \mathbf{R} can also be given by

$$(50.1) \quad \mu^*(x) = \sup_{I \ni x} \frac{\mu(I)}{|I|},$$

where now the supremum is taken over all closed intervals $I = [a, b]$ that contain x and have positive length $|I| = b - a$. The previous definition is clearly less than or equal to this one, since every open interval (a, b) is contained in a closed interval $[a, b]$ with the same length, and

$$(50.2) \quad \mu((a, b)) \leq \mu([a, b]).$$

In the other direction, one can approximate closed intervals by open intervals that contain them.

Let α be a bounded monotone increasing real-valued function on the real line. If μ_α is the corresponding measure, as in Section 44, then its maximal function can be expressed directly in terms of α , by

$$(50.3) \quad \mu_\alpha^*(x) = \sup_{\substack{a \leq x \leq b \\ a < b}} \frac{\alpha(b) - \alpha(a)}{b - a}.$$

More precisely, the supremum is taken over $a, b \in \mathbf{R}$ with $a \leq x \leq b$ and $a < b$, and this expression for the maximal function is trapped between the previous two, by (44.3). If $E_t = \{x \in \mathbf{R} : \mu_\alpha^*(x) > t\}$, then the main estimate from Section 46 can be reformulated as

$$(50.4) \quad |E_t| \leq 2t^{-1} \left(\sup_{x \in \mathbf{R}} \alpha(x) - \inf_{x \in \mathbf{R}} \alpha(x) \right).$$

Now let $(M, d(x, y))$ be a metric space, and let $f : [a, b] \rightarrow M$ be a path of finite length. Let $\alpha(r)$ be the length Λ_a^r of the restriction of f to $[a, r]$ when $a \leq r \leq b$, and put $\alpha(r) = 0$ when $r < a$, $\alpha(r) = \Lambda_a^b$ when $r > b$. Thus α is a bounded monotone increasing function on \mathbf{R} , and

$$(50.5) \quad d(f(r), f(r')) \leq \Lambda_r^{r'} = \alpha(r') - \alpha(r)$$

when $a \leq r \leq r' \leq b$. If $[r, r']$ contains an element of $\mathbf{R} \setminus E_t$, where $t > 0$ and E_t is as in the previous paragraph, then

$$(50.6) \quad d(f(r), f(r')) \leq \alpha(r') - \alpha(r) \leq t(r' - r).$$

In particular, the restriction of f to $[a, b] \setminus E_t$ is Lipschitz with constant t . Note that $[a, b] \setminus E_t$ is a closed set, because E_t is open. Also, (50.4) reduces to

$$(50.7) \quad |E_t| \leq 2t^{-1} \Lambda_a^b.$$

If our metric space is a real or complex vector space with a norm, then we can extend the restriction of f to $[a, b] \setminus E_t$ to a t -Lipschitz function f_t on $[a, b]$. Remember that E_t can be expressed as the union of finitely or countably many pairwise-disjoint open intervals, since E_t is an open set in \mathbf{R} . If I is one of these open intervals and $I \subseteq [a, b]$, then f_t is defined on I as the affine function that agrees with f on the endpoints. If a or b is an element of E_t , and I is an open interval in E_t that contains a or b and whose other endpoint is in $[a, b]$, then we can take f_t to be the constant on $I \cap [a, b]$ that agrees with f at the other endpoint of I . Of course, if $[a, b] \subseteq E_t$, then there is nothing to do.

51 Vector-valued functions

Let V be a real or complex vector space with a norm. As usual, a function $F : [a, b] \rightarrow V$ is said to be differentiable at $x \in (a, b)$ if

$$(51.1) \quad \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h}$$

exists in V . One can also consider one-sided limits at the endpoints.

For example, let V be $L^1([0, 1])$, with respect to Lebesgue measure. Let $F(x)$ be the indicator function of $[0, x]$ as an element of $L^1([0, 1])$ for each $x \in [0, 1]$. It is easy to see that

$$(51.2) \quad \|F(x) - F(y)\|_1 = |x - y|$$

for every $x, y \in [0, 1]$, so that F is actually an isometric embedding of $[0, 1]$ in $L^1([0, 1])$. However, one can also check that F is not differentiable at any point in $[0, 1]$. The derivative of F at $x \in [0, 1]$ is basically a Dirac mass at x , in a weak sense that we shall discuss later.

Now let V be $L^\infty(\mathbf{R})$. If f is a bounded real or complex-valued Lipschitz function on \mathbf{R} , then let $F : \mathbf{R} \rightarrow L^\infty(\mathbf{R})$ be the mapping that sends $x \in \mathbf{R}$ to the translate $f_x(\cdot) = f(\cdot - x)$ of f by x . It is easy to see that this is a Lipschitz mapping from the real line into $L^\infty(\mathbf{R})$, because f is a Lipschitz function on \mathbf{R} . If F is differentiable at any point in \mathbf{R} as a mapping into $L^\infty(\mathbf{R})$, then the difference quotient for f would converge uniformly on \mathbf{R} . This would imply that f is continuously differentiable on \mathbf{R} , with uniformly continuous derivative. Conversely, if f is continuously differentiable on \mathbf{R} , with uniformly continuous derivative, then the difference quotient for f does converge uniformly to the derivative of f , and F is differentiable at every point in \mathbf{R} . More precisely, the derivative of F at $x \in \mathbf{R}$ corresponds to -1 times the derivative of f translated by x in this case. If F is not bounded, then one can take $F(x) = f_x - f$, and get similar conclusions.

Let V be any vector space with a norm $\|v\|$ again, and suppose that F, G are V -valued functions on an interval $[a, b]$ with finite length. One can check that $F - G$ also has finite length on $[a, b]$, which is less than or equal to the sum of the lengths of F and G . It follows that $\|F - G\|$ has finite length as a real-valued function on $[a, b]$, which is to say that it has bounded variation. In particular, $\|F - G\|$ is differentiable almost everywhere as a real-valued function on $[a, b]$. If $x \in [a, b]$ is a limit point of the set where $F = G$, and hence a limit point of the set where $\|F - G\| = 0$, and if $\|F - G\|$ is differentiable at x , then the derivative of $\|F - G\|$ at x is equal to 0. This implies that the derivative of $F - G$ exists at x and is equal to 0, under these conditions. In particular, this can be applied to Lipschitz approximations G of F as in the previous section.

52 Uniform boundedness, 4

Let W be a real or complex vector space with a norm $\|w\|$, and let $\{\lambda_j\}_{j=1}^\infty$ be a sequence of bounded linear functionals on W . Suppose that the dual norms

of the λ_j 's are uniformly bounded, so that

$$(52.1) \quad \|\lambda_j\|_* \leq L$$

for some $L \geq 0$ and each j . Under these conditions, one can check that the set of $w \in W$ such that $\{\lambda_j(w)\}_{j=1}^\infty$ is a Cauchy sequence in \mathbf{R} or \mathbf{C} , as appropriate, is closed. Because of the completeness of the real and complex numbers, this is the same as saying that the set of $w \in W$ such that $\{\lambda_j(w)\}_{j=1}^\infty$ converges in \mathbf{R} or \mathbf{C} is closed. It is easy to see that this is also a linear subspace of W .

In particular, $\{\lambda_j(w)\}_{j=1}^\infty$ converges for every $w \in W$ if it converges for a set of w 's whose linear span is dense in W . In this case,

$$(52.2) \quad \lambda(w) = \lim_{j \rightarrow \infty} \lambda_j(w)$$

defines a linear functional on W . More precisely, λ is a bounded linear functional on W , with

$$(52.3) \quad \|\lambda\|_* \leq L,$$

because of (52.1).

Conversely, if $\{\lambda_j(w)\}_{j=1}^\infty$ converges for every $w \in W$, then $\{\lambda_j(w)\}_{j=1}^\infty$ is bounded for every $w \in W$. The Banach–Steinhaus theorem implies that the λ_j 's have uniformly bounded dual norms when W is complete, as in Section 29.

Suppose now that E is a set of real numbers, and that for each $t \in E$ we have a bounded linear functional λ_t on W . Suppose also that 0 is a limit point of E in \mathbf{R} , and that the λ_t 's have uniformly bounded dual norms. If

$$(52.4) \quad \lim_{\substack{t \rightarrow 0 \\ t \in E}} \lambda_t(w)$$

exists in \mathbf{R} or \mathbf{C} , as appropriate, for a set of $w \in W$ whose linear span is dense in W , then this limit exists for every $w \in W$, and determines a bounded linear functional on W . This is a variant of the earlier discussion for sequences. One can also apply the previous remarks to sequences of elements of E that converge to 0.

53 Weak* derivatives

Let W be a real or complex vector space with a norm $\|w\|$, and let $F(x)$ be a function on a closed interval $[a, b]$ in the real line with values in the dual W^* of W . Thus $F(x)(w)$ is a real or complex-valued function of x on $[a, b]$ for each $w \in W$, as appropriate. If $F(x)$ has finite length as a mapping from $[a, b]$ into W^* , then $F(x)(w)$ has bounded variation as a real or complex-valued function of x on $[a, b]$ for every $w \in W$. This implies that for each $w \in W$ there is a set $Z(w) \subseteq [a, b]$ of Lebesgue measure 0 such that $F(x)(w)$ is differentiable for every $x \in [a, b] \setminus Z(w)$.

Suppose that W is separable, so that there is a collection $\{w_l\}_l$ of finitely or countably many elements of W whose linear span is dense in W . Thus

$Z = \bigcup_l Z(w_l)$ also has Lebesgue measure 0. If $x \in [a, b] \setminus Z$, then $F(x)(w_l)$ is differentiable at x for each l .

We also know that

$$(53.1) \quad \sup_{\substack{a \leq y \leq b \\ y \neq x}} \frac{\|F(x) - F(y)\|_*}{|x - y|} < \infty$$

for almost every $x \in [a, b]$. This follows from the finiteness almost everywhere of the maximal function associated to the function Λ_a^r that measures the length of F on $[a, r]$, as in Section 50. If x has this property and $x \notin Z$, then one can check that the derivative

$$(53.2) \quad \lim_{h \rightarrow 0} \frac{F(x+h)(w) - F(x)(w)}{h}$$

of $F(x)(w)$ at x exists for every $w \in W$, using the remarks in the previous section. Hence the derivative

$$(53.3) \quad \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h}$$

exists for almost every $x \in [a, b]$ in the weak* topology under these conditions.

Let W be the space of continuous real or complex-valued functions on $[0, 1]$ with the supremum norm, so that W^* can be identified with the space of real or complex Borel measures on $[0, 1]$, as appropriate. Also let $F(x)$ be the function on $[0, 1]$ with values in W^* that assigns to $x \in [0, 1]$ the measure on $[0, 1]$ that is Lebesgue measure on $[0, x]$. This is basically the same as the function on $[0, 1]$ with values in $L^1([0, 1])$ discussed in Section 51, by identifying integrable functions on $[0, 1]$ with absolutely continuous measures with respect to Lebesgue measure. Now that we consider F to take values in W^* , it is easy to see that the derivative of F exists with respect to the weak* topology on W^* at every $x \in [0, 1]$, and corresponds to a Dirac mass at x .

54 Lipschitz functions

Let f be a real or complex-valued Lipschitz function on the real line. Thus f is differentiable almost everywhere, since it has bounded variation on any bounded interval. In particular,

$$(54.1) \quad \lim_{j \rightarrow \infty} \frac{f(x+h_j) - f(x)}{h_j} = f'(x)$$

almost everywhere for every sequence $\{h_j\}_{j=1}^\infty$ of nonzero real numbers that converges to 0. This implies that

$$(54.2) \quad \lim_{j \rightarrow \infty} \int_{\mathbf{R}} \frac{f(x+h_j) - f(x)}{h_j} \phi(x) dx = \int_{\mathbf{R}} f'(x) \phi(x) dx$$

for every integrable function ϕ on \mathbf{R} , by the dominated convergence theorem. More precisely, this also uses the fact that the difference quotients are uniformly bounded, because f is Lipschitz. Hence

$$(54.3) \quad \lim_{h \rightarrow 0} \int_{\mathbf{R}} \frac{f(x+h) - f(x)}{h} \phi(x) dx = \int_{\mathbf{R}} f'(x) \phi(x) dx.$$

This is the same as saying that

$$(54.4) \quad \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = f'(x)$$

in the weak* topology on $L^\infty(\mathbf{R})$, as the dual of $L^1(\mathbf{R})$.

Alternatively, we can start with the identity

$$(54.5) \quad \int_{\mathbf{R}} \frac{f(x+h) - f(x)}{h} \phi(x) dx = - \int_{\mathbf{R}} f(x) \frac{\phi(x) - \phi(x-h)}{h} dx,$$

which uses the change of variables $x \mapsto x-h$. This implies that

$$(54.6) \quad \lim_{h \rightarrow 0} \int_{\mathbf{R}} \frac{f(x+h) - f(x)}{h} \phi(x) dx = - \int_{\mathbf{R}} f(x) \phi'(x) dx$$

when ϕ is a continuously-differentiable function with compact support on \mathbf{R} , for instance. Thus

$$(54.7) \quad \lambda_h(\phi) = \int_{\mathbf{R}} \frac{f(x+h) - f(x)}{h} \phi(x) dx$$

defines a bounded family of linear functionals on $L^1(\mathbf{R})$ that converges as $h \rightarrow 0$ on a dense linear subspace of $L^1(\mathbf{R})$, and hence converges on all of $L^1(\mathbf{R})$, as in Section 52. The limit is a bounded linear functional on $L^1(\mathbf{R})$ that can be expressed by integration with an element of $L^\infty(\mathbf{R})$, that corresponds to the derivative of f .

If f is a bounded Lipschitz function on \mathbf{R} , then we can take $F : \mathbf{R} \rightarrow L^\infty(\mathbf{R})$ to be the function that sends a real number to the corresponding translate of f , as in Section 51. Otherwise, we can take a difference between f and its translate to get an element of $L^\infty(\mathbf{R})$, as before. This defines a Lipschitz mapping from \mathbf{R} into $L^\infty(\mathbf{R})$, with a weak* derivative at every point.

55 Averages

Let f be a locally integrable function on the real line, and put

$$(55.1) \quad A_h(f)(x) = \frac{1}{h} \int_x^{x+h} f(y) dy$$

for every $h, x \in \mathbf{R}$ with $h \neq 0$. As before, the integral in this expression is considered to be oriented, as in ordinary calculus, so that

$$(55.2) \quad A_h(f)(x) = \frac{1}{|h|} \int_{x-|h|}^x f(y) dy$$

when $h < 0$. In particular,

$$(55.3) \quad |A_h(f)(x)| \leq A_h(|f|)(x).$$

If $f \in L^p(\mathbf{R})$, $1 \leq p \leq \infty$, then $A_h(f) \in L^p(\mathbf{R})$ for every $h \neq 0$, and

$$(55.4) \quad \|A_h(f)\|_p \leq \|f\|_p.$$

This is very easy to see when $p = \infty$. If $p = 1$, then one can integrate (55.3) in x , and then use Fubini's theorem. If $1 < p < \infty$, then

$$(55.5) \quad |A_h(f)(x)|^p \leq A_h(|f|^p)(x),$$

by the convexity of r^p on the nonnegative real numbers, as in Jensen's inequality. One can then integrate in x and apply Fubini's theorem, as when $p = 1$.

If f is continuous at x , then

$$(55.6) \quad \lim_{h \rightarrow 0} A_h(f)(x) = f(x).$$

If f is uniformly continuous, then this holds with uniform convergence. If f is a continuous function on \mathbf{R} , then f is uniformly continuous on bounded sets, and we get uniform convergence on bounded sets.

If $f \in L^p(\mathbf{R})$, $1 \leq p < \infty$, then

$$(55.7) \quad \lim_{h \rightarrow 0} \|A_h(f) - f\|_p = 0.$$

To see this, observe first that this holds for every continuous function f with compact support on the real line. More precisely, f is uniformly continuous in this case, so that $A_h(f)$ converges to f uniformly as $h \rightarrow 0$, as in the previous paragraph. Also, the support of $A_h(f)$ is contained in a single compact set when $|h| \leq 1$, say, and hence uniform convergence implies convergence in the $L^p(\mathbf{R})$ norm. Any $f \in L^p(\mathbf{R})$ can be approximated in the L^p norm by a continuous function with compact support when $p < \infty$, and one can get (55.7) using this approximation and the uniform bounds for A_h on $L^p(\mathbf{R})$.

56 L^p derivatives

If f, g are locally integrable functions on the real line, then we say that $f' = g$ in the sense of distributions if

$$(56.1) \quad \int_{\mathbf{R}} f(x) \phi'(x) dx = - \int_{\mathbf{R}} g(x) \phi(x) dx$$

for every continuously-differentiable function ϕ with compact support on \mathbf{R} . If f is continuously differentiable on \mathbf{R} , then the ordinary derivative of f has this property, by integration by parts. Similarly, if

$$(56.2) \quad \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = g(x)$$

with respect to the L^1 norm on any bounded interval in the real line, then $f' = g$ in the sense of distributions. This follows from (54.5), by taking the limit as $h \rightarrow 0$.

Suppose that

$$(56.3) \quad f(x+h) - f(x) \in L^p(\mathbf{R})$$

for some p , $1 \leq p < \infty$, and every $h \in \mathbf{R}$, which holds in particular when $f \in L^p(\mathbf{R})$. We say that f is differentiable in the L^p sense, with derivative equal to g , if $g \in L^p(\mathbf{R})$, and one has convergence in (56.2) in the L^p norm. This implies that $f' = g$ in the sense of distributions, as in the previous paragraph. If $g \in L^p(\mathbf{R})$ and

$$(56.4) \quad f(x) = \int_a^x g(y) dy$$

for some $a \in \mathbf{R}$, then

$$(56.5) \quad \frac{f(x+h) - f(x)}{h} = A_h(g)(x)$$

converges to g as $h \rightarrow 0$ in the L^p norm, as in the previous section, and so the derivative of f is equal to g in the L^p sense. If g is locally integrable, then $A_h(g) \rightarrow g$ as $h \rightarrow 0$ in the L^1 norm on every bounded interval, and we still have that $f' = g$ in the sense of distributions.

Note that $f' = 0$ in the sense of distributions when

$$(56.6) \quad \int_{\mathbf{R}} f(x) \phi'(x) dx = 0$$

for every continuously-differentiable function ϕ with compact support. If ψ is a continuous function with compact support on \mathbf{R} such that

$$(56.7) \quad \int_{\mathbf{R}} \psi(y) dy = 0,$$

then

$$(56.8) \quad \phi(x) = \int_{-\infty}^x \psi(y) dy$$

is continuously differentiable and has compact support, and $\phi' = \psi$. Thus $f' = 0$ in the sense of distributions if and only if

$$(56.9) \quad \int_{\mathbf{R}} f(x) \psi(x) dx = 0$$

for every continuous function ψ with compact support and integral 0. One can show that this happens if and only if f is constant almost everywhere.

If $f' = g$ in the sense of distributions, then it follows that the difference between f and (56.4) is constant almost everywhere, since they have the same derivative. In particular, (56.5) holds for each $h \neq 0$ and almost every x . If $g \in L^p(\mathbf{R})$, then we get that the derivative of f is equal to g in the L^p sense, as before.

57 L^p Lipschitz conditions

Let f be a locally integrable function on the real line that satisfies (56.3) for some p , $1 \leq p < \infty$, and every $h \in \mathbf{R}$, such as an L^p function. Suppose that

$$(57.1) \quad \left(\int_{\mathbf{R}} |f(x+h) - f(x)|^p dx \right)^{1/p} \leq C|h|$$

for some $C \geq 0$ and every $h \in \mathbf{R}$, which is the same as saying that

$$(57.2) \quad \frac{f(x+h) - f(x)}{h}$$

is uniformly bounded in $L^p(\mathbf{R})$. Note that this happens when $f' = g \in L^p(\mathbf{R})$ in the sense of distributions, since the difference quotient is equal to $A_h(g)$.

Suppose also that $1 < p < \infty$, and let q be the conjugate exponent to p , $1/p + 1/q = 1$. If λ_h is as in (54.7) for $h \neq 0$, then λ_h is a uniformly bounded family of linear functionals on $L^q(\mathbf{R})$, by Hölder's inequality. As in Section 54,

$$(57.3) \quad \lim_{h \rightarrow 0} \lambda_h(\phi) = - \int_{\mathbf{R}} f(x) \phi'(x) dx$$

for every continuously-differentiable function ϕ with compact support on \mathbf{R} . Because these functions are dense in $L^q(\mathbf{R})$, it follows that

$$(57.4) \quad \lim_{h \rightarrow 0} \lambda_h(\phi)$$

exists for every $\phi \in L^q(\mathbf{R})$, as in Section 52. The limit determines a bounded linear functional on $L^q(\mathbf{R})$, and so there is a function $g \in L^p(\mathbf{R})$ such that

$$(57.5) \quad \lim_{h \rightarrow 0} \lambda_h(\phi) = \int_{\mathbf{R}} g(x) \phi(x) dx$$

for every $\phi \in L^q(\mathbf{R})$. In particular, this holds when ϕ is a continuously-differentiable function with compact support on \mathbf{R} , for which we have (57.3). This shows that $f' = g$ in the sense of distributions.

If $p = 1$, then it is better to think of λ_h as a uniformly bounded family of linear functionals on the space $C_0(\mathbf{R})$ of continuous functions on the real line that vanish at infinity, equipped with the supremum norm. We still have (57.3) for every continuously-differentiable function ϕ with compact support on \mathbf{R} , and hence that (57.4) exists for every $\phi \in C_0(\mathbf{R})$, as in Section 52. The limit determines a bounded linear functional on $C_0(\mathbf{R})$, and so there is a real or complex Borel measure μ on \mathbf{R} such that

$$(57.6) \quad \lim_{h \rightarrow 0} \lambda_h(\phi) = \int_{\mathbf{R}} \phi d\mu$$

for every $\phi \in C_0(\mathbf{R})$. Combining this with (57.3), we get that

$$(57.7) \quad \int_{\mathbf{R}} f(x) \phi'(x) dx = - \int_{\mathbf{R}} \phi d\mu$$

for every continuously-differentiable function ϕ with compact support on \mathbf{R} . This can be expressed by saying that $f' = \mu$ in the sense of distributions.

If α is a function of bounded variation on \mathbf{R} , and if μ_α is the corresponding real or complex Borel measure as in Section 44, then $\alpha' = \mu$ in the sense of distributions. This is basically another version of integration by parts. One can also show that every real or complex Borel measure on the real line is of this form. If f is a locally integrable function on \mathbf{R} such that $f' = \mu$ in the sense of distributions for some real or complex Borel measure μ , then it follows that f is equal almost everywhere to a function of bounded variation. Conversely, one can check that such functions satisfy the integrated Lipschitz condition (57.1) with $p = 1$.

58 Dyadic intervals

In this section, it will be convenient to use $[0, 1)$ as the unit interval, consisting of $x \in \mathbf{R}$ with $0 \leq x < 1$. By a *dyadic subinterval* of the unit interval we mean an interval of the form $[j 2^{-l}, (j+1) 2^{-l})$, where j, l are nonnegative integers and $j < 2^l$. Thus the unit interval is the disjoint union of these dyadic intervals at level l . If I, I' are dyadic intervals of arbitrary lengths, then either $I \subseteq I'$, $I' \subseteq I$, or $I \cap I' = \emptyset$. More precisely, if $|I| \leq |I'|$, where $|I|$ denotes the length of I , then either $I \subseteq I'$ or $I \cap I' = \emptyset$.

Let μ be a positive Borel measure on $[0, 1)$. The *dyadic maximal function* associated to μ is defined by

$$(58.1) \quad \mu_\delta^*(x) = \sup_{x \in I} \frac{\mu(I)}{|I|},$$

where now the supremum is taken over all dyadic intervals that contain a given point $x \in [0, 1)$. Similarly, if f is an integrable function on $[0, 1)$, then we put

$$(58.2) \quad f_\delta^*(x) = \sup_{x \in I} \frac{1}{|I|} \int_I |f(y)| dy,$$

where again the supremum is taken over all dyadic intervals I such that $x \in I$. This is the same as $\mu_\delta^*(x)$, where μ is the Borel measure on $[0, 1)$ defined by

$$(58.3) \quad \mu(A) = \int_A |f(y)| dy,$$

as in Section 46.

Consider

$$(58.4) \quad E_{\delta,t} = \{x \in [0, 1) : \mu_\delta^*(x) > t\}$$

for each $t > 0$. Thus $x \in E_{\delta,t}$ if and only if there is a dyadic interval I such that $x \in I$ and

$$(58.5) \quad \mu(I) > t |I|,$$

in which case $I \subseteq E_{\delta,t}$. Let $I(x)$ be the maximal dyadic interval that contains x and satisfies (58.5) for each $x \in E_{\delta,t}$. If $x, y \in E_{\delta,t}$, then either $I(x) = I(y)$

or $I(x) \cap I(y) = \emptyset$, by maximality and the nesting properties of dyadic intervals mentioned before.

Let \mathcal{M}_t be the collection of dyadic intervals of the form $I(x)$ for some x in $E_{\delta,t}$. Note that the elements of \mathcal{M}_t are pairwise disjoint, and

$$(58.6) \quad \bigcup_{I \in \mathcal{M}_t} I = E_{\delta,t}.$$

Hence

$$(58.7) \quad |E_{\delta,t}| = \sum_{I \in \mathcal{M}_t} |I| < t^{-1} \sum_{I \in \mathcal{M}_t} \mu(I) = t^{-1} \mu(E_{\delta,t}).$$

This is almost the same as the estimate in Section 46, but without the additional factor of 2. Although we have focused on dyadic subintervals of the unit interval for simplicity, there is an analogous discussion for arbitrary dyadic intervals in the real line, and the corresponding maximal functions.

59 Dyadic averages

Let f be an integrable function on $[0, 1)$, and put

$$(59.1) \quad A_l(f)(x) = 2^l \int_{j 2^{-l}}^{(j+1) 2^{-l}} f(y) dy$$

when $j 2^{-l} \leq x < (j+1) 2^{-l}$. Thus $A_l(f)(x)$ is the average of f over the dyadic interval of length 2^{-l} that contains x . In particular, $A_l(f)$ is constant on dyadic intervals of length 2^{-l} , by construction. Also,

$$(59.2) \quad \begin{aligned} \int_0^1 A_l(f)(x) dx &= \sum_{j=0}^{2^l-1} \int_{j 2^{-l}}^{(j+1) 2^{-l}} A_l(f)(x) dx \\ &= \sum_{j=0}^{2^l-1} \int_{j 2^{-l}}^{(j+1) 2^{-l}} f(x) dx = \int_0^1 f(x) dx. \end{aligned}$$

If $f \in L^p([0, 1))$, $1 \leq p \leq \infty$, then

$$(59.3) \quad \|A_l(f)\|_p \leq \|f\|_p.$$

This is immediate when $p = \infty$. Note that

$$(59.4) \quad |A_l(f)(x)| \leq A_l(|f|)(x)$$

for every $x \in [0, 1)$ and $l \geq 0$, and that

$$(59.5) \quad |A_l(f)(x)|^p \leq A_l(|f|^p)(x)$$

when $f \in L^p([0, 1))$, $1 < p < \infty$, by Jensen's inequality. To estimate $\|A_l(f)\|_p$, one can integrate these inequalities using the identity in the previous paragraph.

As in Section 55,

$$(59.6) \quad \lim_{l \rightarrow \infty} A_l(f)(x) = f(x)$$

when f is continuous at x , and with uniform convergence when f is uniformly continuous on $[0, 1]$. If f is a continuous function on $[0, 1]$, then f is uniformly continuous, by compactness. If $f \in L^p([0, 1])$, $1 \leq p < \infty$, then

$$(59.7) \quad \lim_{l \rightarrow \infty} \|A_l(f) - f\|_p = 0.$$

This follows from uniform convergence when f is a continuous function on $[0, 1]$, and otherwise one can approximate by continuous functions using the uniform bound (59.3). Of course, $L^p([0, 1])$ is the same as $L^p([0, 1])$, and so continuous functions on $[0, 1]$ are still dense in this space when $p < \infty$.

If $f \in L^1([0, 1])$, then Lebesgue's theorem implies that (59.6) holds almost everywhere on $[0, 1]$. More precisely,

$$(59.8) \quad \lim_{l \rightarrow \infty} 2^l \int_{I_l(x)} |f(y) - f(x)| dy = 0$$

for almost every $x \in [0, 1]$, where $I_l(x)$ denotes the dyadic interval of length 2^{-l} that contains x . This follows from Lebesgue's theorem, as in Section 47, and one can also establish it a bit more directly. Specifically, one can use the estimate for the dyadic maximal function in the previous section, instead of the estimate for the Hardy–Littlewood maximal function in Section 46.

60 Rademacher functions

Let $r_1(x), r_2(x), \dots$ be the functions defined on $[0, 1]$ by

$$(60.1) \quad \begin{aligned} r_l(x) &= 1 && \text{when } j 2^{-l} \leq x < (j+1) 2^{-l} \text{ and } j \text{ is even} \\ &= -1 && \text{when } j 2^{-l} \leq x < (j+1) 2^{-l} \text{ and } j \text{ is odd.} \end{aligned}$$

Thus $r_l(x)$ is constant on each dyadic interval of length 2^{-l} ,

$$(60.2) \quad \int_I r_l(x) dx = 0$$

for each dyadic interval I of length 2^{-l+1} , and $|r_l(x)| = 1$ for every $x \in [0, 1]$ and positive integer l . These are known as the *Rademacher functions* on the unit interval.

Let X be the set of sequences $x = \{x_k\}_{k=1}^{\infty}$ with $x_k = 1$ or -1 for each k . Equivalently, X is the Cartesian product of a sequence of copies of $\{1, -1\}$. This is a compact Hausdorff topological space with respect to the product topology, which is homeomorphic to the usual middle-thirds Cantor set. There is a natural continuous mapping from X onto the closed unit interval $[0, 1]$, defined by

$$(60.3) \quad \beta(x) = \sum_{k=1}^{\infty} \left(\frac{x_k + 1}{2} \right) 2^{-k}.$$

Each element x of X corresponds to an infinite binary sequence $\{(x_k + 1)/2\}_{k=1}^\infty$, and β sends x to the real number with that binary expansion. Every real number in $[0, 1]$ has a binary expansion, and the binary expansion is unique for all but a countable set of real numbers. Dyadic rational numbers of the form $j 2^{-l}$, $0 < j < 2^l$, have two binary expansions, which agree up to a point where one has a 1 followed by all 0's, and the other has a 0 followed by all 1's.

There is a natural Borel probability measure on X , which is the product measure associated to 1, -1 having probability $1/2$ in each coordinate. This probability measure corresponds exactly to Lebesgue measure on $[0, 1]$ under the mapping β . That β fails to be one-to-one on a countable set does not really matter here, since countable sets have measure 0. Thus $[0, 1]$ and X are basically the same as probability spaces. The Rademacher functions r_l on $[0, 1]$ correspond to the coordinate functions $x \mapsto x_l$ on X , which are independent identically distributed random variables.

In particular,

$$(60.4) \quad \int_0^1 r_{l_1}(x) r_{l_2}(x) \cdots r_{l_n}(x) dx = 0$$

when $1 \leq l_1 < l_2 < \cdots < l_n$. Because of independence, the integral of the product should be the same as the product of the individual integrals, each of which is 0, by (60.2). One can see this more directly by observing that the integral over each dyadic interval of length 2^{-l_n+1} is 0, because the integral of r_{l_n} over such an interval is 0, as in (60.2), while the other functions in the integral are constant over these intervals.

61 L^p estimates

The Rademacher functions are orthonormal in $L^2([0, 1])$, since

$$(61.1) \quad \|r_l\|_2 = \left(\int_0^1 |r_l(x)|^2 dx \right)^{1/2} = 1$$

for each l , and

$$(61.2) \quad \langle r_k, r_l \rangle = \int_0^1 r_k(x) r_l(x) dx = 0$$

when $k \neq l$. This implies that

$$(61.3) \quad \left\| \sum_{l=1}^n a_l r_l \right\|_2 = \left(\sum_{l=1}^n a_l^2 \right)^{1/2}$$

for every $a_1, \dots, a_n \in \mathbf{R}$. Let us check that

$$(61.4) \quad \left\| \sum_{l=1}^n a_l r_l \right\|_\infty = \sum_{l=1}^n |a_l|.$$

The left side of (61.4) is clearly less than or equal to the right side, by the triangle inequality. To get the opposite inequality, one can choose a dyadic interval of length 2^{-n} on which $a_l r_l = |a_l|$ for $l = 1, \dots, n$.

Before proceeding, it will be helpful to remember two basic facts about L^p norms. The first is that

$$(61.5) \quad \|f\|_p = \left(\int_0^1 |f(x)|^p dx \right)^{1/p}$$

is monotone increasing in p , by Jensen's inequality. The second fact is that the L^p norm is logarithmically convex in $1/p$, which means that

$$(61.6) \quad \|f\|_r \leq \|f\|_p^t \|f\|_q^{1-t}$$

when $p, q, r > 0$, $0 < t < 1$, and

$$(61.7) \quad \frac{1}{r} = \frac{t}{p} + \frac{1-t}{q}.$$

This can be derived from Hölder's inequality. It is a little simpler to start with the $r = 1$ case, and then get (61.6) by applying the $r = 1$ case to $|f|^r$.

If $2 < p < \infty$, then there is a constant $C(p) > 0$ such that

$$(61.8) \quad \left\| \sum_{l=1}^n a_l r_l \right\|_p \leq C(p) \left(\sum_{l=1}^n a_l^2 \right)^{1/2}$$

for every $a_1, \dots, a_n \in \mathbf{R}$. Of course, it is very important here that $C(p)$ does not depend on n . To prove (61.8), it suffices to restrict our attention to $p = 2^k$ for some positive integer $k \geq 2$, because of the monotonicity of the L^p norm. One can get better constants for the intermediate exponents using (61.3) and (61.6). If $p = 2^k$, then one can expand

$$(61.9) \quad \left\| \sum_{l=1}^n a_l r_l \right\|_{2^k}^{2^k} = \int_0^1 \left(\sum_{l=1}^n a_l r_l \right)^{2^k} dx$$

into a 2^k -fold sum, where each term has the product of 2^k coefficients a_l times the integral of the product of 2^k Rademacher functions r_l . As in the previous section, most of these integrals are equal to 0. The only way that the integral is not equal to 0 is to have r_l occur an even number of times for each l . In this case, the integral is equal to 1, and the coefficients are products of 2^{k-1} factors of r_l^2 , $1 \leq l \leq n$. This permits one to estimate the 2^k -fold sum by a constant multiple of

$$(61.10) \quad \left(\sum_{l=1}^n a_l^2 \right)^{2^{k-1}},$$

as desired. The $k = 2$ case is already a nice exercise.

If $0 < p < 2$, then there is a constant $C(p) > 0$ such that

$$(61.11) \quad \left(\sum_{l=1}^n a_l^2 \right)^{1/2} \leq C(p) \left\| \sum_{l=1}^n a_l r_l \right\|_p$$

for every $a_1, \dots, a_n \in \mathbf{R}$. Again, it is very important that $C(p)$ not depend on n . This time, we can apply (61.6) to $f = \sum_{l=1}^n a_l r_l$, $r = 2$, and $q = 4$ to get that

$$(61.12) \quad \left(\sum_{l=1}^n a_l^2 \right)^{1/2} \leq \left\| \sum_{l=1}^n a_l r_l \right\|_p^t \left\| \sum_{l=1}^n a_l r_l \right\|_4^{1-t}$$

for some t , $0 < t < 1$. Using the previous estimate with $p = 4$, we get that

$$(61.13) \quad \left(\sum_{l=1}^n a_l^2 \right)^{1/2} \leq C(4) \left(\sum_{l=1}^n a_l^2 \right)^{(1-t)/2} \left\| \sum_{l=1}^n a_l r_l \right\|_p^t.$$

This implies (61.11), by dividing both sides by $\left(\sum_{l=1}^n a_l^2 \right)^{(1-t)/2}$, at least when $a_l \neq 0$ for some l .

62 Rademacher sums

Let a_1, a_2, \dots be a sequence of real numbers such that $\sum_{l=1}^{\infty} a_l^2$ converges, and consider

$$(62.1) \quad f(x) = \sum_{l=1}^{\infty} a_l r_l(x).$$

This series converges in $L^2([0, 1])$, by the orthonormality of the Rademacher functions. Moreover, the series converges in $L^p([0, 1])$ for every $p < \infty$, by the estimates in the previous section. Using these estimates, one can also check that this series converges in $L^p([0, 1])$ in the generalized sense for every $p < \infty$, as in Section 14.

Observe that

$$(62.2) \quad A_n(f)(x) = \sum_{l=1}^n a_l r_l(x)$$

for every n , where A_n is the dyadic averaging operator in Section 59. This follows from the fact that $A_n(r_l) = 0$ when $l > n$. By Lebesgue's theorem,

$$(62.3) \quad \lim_{n \rightarrow \infty} A_n(f)(x) = f(x)$$

almost everywhere on $[0, 1]$, which implies that the series defining f converges almost everywhere. However, if $\sum_{l=1}^{\infty} a_l r_l(x)$ converges in the generalized sense as a sum of real numbers for any $x \in [0, 1]$, then

$$(62.4) \quad \sum_{l=1}^{\infty} |a_l r_l(x)| = \sum_{l=1}^{\infty} |a_l|$$

converges, as in Section 3. Similarly, if $f \in L^\infty([0, 1))$, then $A_n(f)$ is uniformly bounded, and hence $\sum_{l=1}^\infty |a_l|$ converges, by (61.4).

Let π be a one-to-one mapping from the set \mathbf{Z}_+ of positive integers onto itself, and let X be the space of all sequences $\{x_k\}_{k=1}^\infty$ with $x_k = \pm 1$ for each k , as in Section 60. Thus π determines a measure-preserving homeomorphism from X onto itself, which sends $\{x_k\}_{k=1}^\infty$ to $\{x_{\pi(k)}\}_{k=1}^\infty$. Using this transformation, one can check that $\sum_{l=1}^\infty a_{\pi(l)} r_{\pi(l)}(x)$ also converges almost everywhere. More precisely, this rearrangement of the series corresponds to the composition of $\sum_{l=1}^\infty a_{\pi(l)} r_l(x)$ with the automorphism on X just mentioned. This new series is of the same type as the previous one, and so converges almost everywhere for the same reasons as before.

63 Lacunary series

Let \mathbf{T} be the unit circle in the complex plane, consisting of $z \in \mathbf{C}$ with $|z| = 1$. It is well known that

$$(63.1) \quad \int_{\mathbf{T}} z^j |dz| = 0$$

for every nonzero integer j , where $|dz|$ denotes the element of arc length. If $j = 0$, then z^j is interpreted as being equal to 1, and the integral is equal to 2π , the circumference of the circle. The usual integral inner product for complex-valued functions in $L^2(\mathbf{T})$ is defined by

$$(63.2) \quad \langle f, g \rangle = \frac{1}{2\pi} \int_{\mathbf{T}} f(z) \overline{g(z)} |dz|,$$

and the corresponding norm is given by

$$(63.3) \quad \|f\|_2 = \left(\frac{1}{2\pi} \int_{\mathbf{T}} |f(z)|^2 |dz| \right)^{1/2}.$$

The functions z^j , $j \in \mathbf{Z}$, are orthonormal with respect to this inner product, because of (63.1) and the fact that the integral is equal to 2π when $j = 0$. It is well known that the linear span of these functions is dense in $L^2(\mathbf{T})$, and more precisely that their linear span is dense in the space of continuous functions on \mathbf{T} with respect to the supremum norm. This implies that z^j , $j \in \mathbf{Z}$, is an orthonormal basis for $L^2(\mathbf{T})$.

Let $n_1 < n_2 < \dots$ be a strictly increasing sequence of positive integers, and let a_1, a_2, \dots be a sequence of complex numbers such that $\sum_{j=1}^\infty |a_j|^2$ converges. Thus

$$(63.4) \quad f(z) = \sum_{j=1}^\infty a_j z^{n_j}$$

converges in $L^2(\mathbf{T})$, since the z^{n_j} 's are orthonormal in $L^2(\mathbf{T})$. We say that (63.4) is a *lacunary* or *gap series* if there is a $q > 1$ such that

$$(63.5) \quad n_{j+1} \geq q n_j$$

for each j . In this case, (63.4) actually converges in $L^p(\mathbf{T})$ for each $p < \infty$. One can also show that the series converges in the generalized sense in $L^p(\mathbf{T})$, as in Section 14.

To see this, it suffices to show that for each $p \in (2, \infty)$ there is a constant $C'(p) > 0$ such that

$$(63.6) \quad \left\| \sum_{j=1}^L a_j z^{n_j} \right\|_p \leq C'(p) \left(\sum_{j=1}^L |a_j|^2 \right)^{1/2}$$

for every $a_1, \dots, a_L \in \mathbf{C}$ and $L \geq 1$. It is also enough to do this when $p = 2^k$ for some integer $k \geq 2$. In this case, the p th power of the L^p norm can be expanded into a 2^k -fold sum, as before. More precisely,

$$(63.7) \quad \left| \sum_{j=1}^L a_j z^{n_j} \right|^{2^k} = \left(\sum_{j=1}^L a_j z^{n_j} \right)^{2^{k-1}} \left(\sum_{j=1}^L \bar{a}_j \bar{z}^{n_j} \right)^{2^{k-1}},$$

since $|a|^2 = a\bar{a}$ for every $a \in \mathbf{C}$. Thus each term in the 2^k -fold sum has 2^{k-1} a_j 's and z^{n_j} 's, and 2^{k-1} \bar{a}_j 's and \bar{z}^{n_j} 's.

Each term is also integrated over \mathbf{T} , and so includes an expression of the form

$$(63.8) \quad \int_{\mathbf{T}} \left(\prod_{l=1}^{2^{k-1}} z^{n_{j_l}} \right) \left(\prod_{l'=1}^{2^{k-1}} \bar{z}^{n_{j'_l}} \right) |dz|,$$

where the j_l 's and $j'_{l'}$'s are integers between 1 and L . Because of (63.1), this integral is equal to 0 unless

$$(63.9) \quad \sum_{l=1}^{2^{k-1}} n_{j_l} - \sum_{l'=1}^{2^{k-1}} n_{j'_{l'}} = 0.$$

If q is large enough, depending on k , then the only way that this can happen is if the largest of the n_{j_l} 's is equal to the largest of the $n_{j'_{l'}}$'s. One can then repeat the argument to get that the n_{j_l} 's and $n_{j'_{l'}}$'s are permutations of each other.

This permits the 2^k -fold sum to be estimated in terms of $\left(\sum_{j=1}^L |a_j|^2 \right)^{2^{k-1}}$, as in Section 61. If q is not sufficiently large for this argument, then one can express (63.4) as a sum of finitely many lacunary series with larger gaps. More precisely, (63.4) can be expressed as the sum of r lacunary series with gaps of size q^r for each positive integer r , by taking every r th term in the series.

64 Walsh functions

If $I = \{l_1, \dots, l_n\}$ is a finite set of positive integers, then the corresponding *Walsh function* w_I on $[0, 1)$ is defined by

$$(64.1) \quad w_I(x) = r_{l_1}(x) r_{l_2}(x) \cdots r_{l_n}(x),$$

where the r_I 's are Rademacher functions. If $I = \emptyset$, then we take w_I to be the constant function 1. Thus

$$(64.2) \quad |w_I(x)| = 1$$

for every $x \in [0, 1)$ and finite set I of positive integers, and

$$(64.3) \quad \int_0^1 w_I(x) dx = 0$$

when $I \neq \emptyset$, as in Section 60. This implies that

$$(64.4) \quad \int_0^1 w_I(x) w_{I'}(x) dx = 0$$

when $I \neq I'$, so that the Walsh functions are orthonormal in $L^2([0, 1))$.

The Walsh functions actually form an orthonormal basis for $L^2([0, 1))$. To see this, it suffices to show that the linear span of the Walsh functions is dense in $L^2([0, 1))$. Note that $w_I(x)$ is constant on dyadic intervals of length 2^{-n} when $I \subseteq \{1, \dots, n\}$, because of the corresponding property of the Rademacher functions. One can check that the linear span of the Walsh functions w_I with $I \subseteq \{1, \dots, n\}$ is exactly the same as the space of functions on $[0, 1)$ that are constant on dyadic intervals of length 2^{-n} . Both spaces have dimension 2^n , for instance, since there are 2^n subsets of $\{1, \dots, n\}$, and 2^n dyadic intervals of length 2^{-n} . It follows that the linear span of all Walsh functions is the space of dyadic step functions on $[0, 1)$, which are the functions that are constant on dyadic intervals of length 2^{-n} for some n . Hence the Walsh functions form an orthonormal basis of $L^2([0, 1))$, because the dyadic step functions are dense in $L^2([0, 1))$.

There is another description of the Walsh functions in terms of harmonic analysis. Let X be the space of sequences $\{x_k\}_{k=1}^\infty$ with $x_k = \pm 1$ for each k , as in Section 60. It is easy to see that X is a commutative group with respect to coordinatewise multiplication. More precisely, X is a topological group with respect to the product topology, because the group operations are continuous with respect to this topology. Note that the probability measure on X described before is invariant under translations defined by this group structure, and hence corresponds to Haar measure on X . The Rademacher functions may be identified with the coordinate functions on X , and so the Walsh functions may be identified with products of coordinate functions on X . One can check that these are continuous homomorphisms from X into the multiplicative group of nonzero complex numbers, and that every such homomorphism arises in this way.

65 Independent random variables

Let $(X_1, \mu_1), \dots, (X_n, \mu_n)$ be probability spaces, and let $X = X_1 \times \dots \times X_n$ be their product, with the product measure $\mu = \mu_1 \times \dots \times \mu_n$. Also let f_1, \dots, f_n

be real or complex-valued functions on X_1, \dots, X_n , respectively, which can be identified with functions on X that are constant in the other variables. Suppose that $f_j \in L^2(X_j, \mu_j)$,

$$(65.1) \quad \int_{X_j} f_j d\mu_j = 0,$$

and

$$(65.2) \quad \|f_j\|_{L^2(X_j, \mu_j)} = \left(\int_{X_j} |f_j|^2 d\mu_j \right)^{1/2} = 1$$

for each j . It may be that the (X_j, μ_j) 's are copies of the same space, for instance, and that the f_j 's are copies of the same function on this space. As functions on X , it is easy to see that f_1, \dots, f_n are orthonormal in $L^2(X, \mu)$. This is because

$$(65.3) \quad \int_X f_j f_l d\mu = \left(\int_{X_j} f_j d\mu_j \right) \left(\int_{X_l} f_l d\mu_l \right) = 0$$

when $j \neq l$ in the real case, and

$$(65.4) \quad \int_X f_j \overline{f_l} d\mu = \left(\int_{X_j} f_j d\mu_j \right) \left(\int_{X_l} \overline{f_l} d\mu_l \right) = 0$$

in the complex case. Hence

$$(65.5) \quad \left\| \sum_{j=1}^n a_j f_j \right\|_{L^2(X, \mu)} = \left(\sum_{j=1}^n |a_j|^2 \right)^{1/2}$$

for any real or complex numbers a_1, \dots, a_n , as appropriate.

Let k be a positive integer, and put $p = 2^k$. Suppose in addition that $f_j \in L^p(X_j, \mu_j)$ for each j , and that

$$(65.6) \quad \|f_j\|_{L^p(X_j, \mu_j)} = \left(\int_{X_j} |f_j|^p d\mu_j \right)^{1/p} \leq L_p$$

for some $L_p \geq 0$ and $j = 1, \dots, n$. In this case, one can show that

$$(65.7) \quad \left\| \sum_{j=1}^n a_j f_j \right\|_{L^p(X, \mu)} \leq C(p, L_p) \left(\sum_{j=1}^n |a_j|^2 \right)^{1/2}$$

for some constant $C(p, L_p) \geq 0$ and all $a_1, \dots, a_n \in \mathbf{R}$ or \mathbf{C} , as appropriate. As usual, it is very important that $C(p, L_p)$ does not depend on n here. To see this, one can expand

$$(65.8) \quad \left\| \sum_{j=1}^n a_j f_j \right\|_{L^p(X, \mu)}^p = \int_X \left| \sum_{j=1}^n a_j f_j \right|^p d\mu$$

into a 2^k -fold sum, where each term is a product of 2^k a_j 's and perhaps their complex conjugates times the integral of a product of 2^k f_j 's and perhaps their

complex conjugates, as in Sections 61 and 63. The integrals can be estimated individually using Hölder's inequality and the hypothesis that the f_j 's have bounded L^p norms. The main point is that the integral is equal to 0 whenever an f_j occurs exactly once for some j , because the integral over X of a product of f_j 's and perhaps their complex conjugates is equal to the product of the integrals over the X_j 's of the corresponding f_j 's for $j = 1, \dots, n$. In the remaining terms, there is a product of 2^k a_j 's and perhaps their complex conjugates, in which each a_j either does not occur or occurs more than once. This permits one to estimate the sum by a constant multiple of

$$(65.9) \quad \left(\sum_{j=1}^n |a_j|^2 \right)^{2^{k-1}},$$

as before. This is a bit more complicated than in the context of Rademacher functions, where the integrals are equal to 0 when any f_j occurs an odd number of times. However, one can use the monotonicity of ℓ^p norms as in Section 9 to deal with this.

These estimates for $p = 2^k$ imply analogous estimates for $2 \leq p \leq 2^k$, as in Section 61. In particular, there are analogous estimates for every $p \in (2, \infty)$ when the f_j 's have bounded L^p norms for each $p \in (2, \infty)$. Using the upper bound for $k = 2$, one also gets that

$$(65.10) \quad \left(\sum_{j=1}^n |a_j|^2 \right)^{1/2} \leq C(p, L_4) \left\| \sum_{j=1}^n a_j f_j \right\|_{L^p(X, \mu)}$$

for $0 < p < 2$, as in Section 61. Here $C(p, L_4)$ is a positive constant that does not depend on n , but does depend on p and the upper bound L_4 for the L^4 norms of the f_j 's.

Suppose now that $(X_1, \mu_1), (X_2, \mu_2), \dots$ is an infinite sequence of probability spaces, $X = \prod_{j=1}^{\infty} X_j$ is their product, and μ is the corresponding product measure on X . Let f_1, f_2, \dots be real or complex-valued functions on X_1, X_2, \dots , respectively, which can be identified with functions on X that are constant in the other variables. As before, suppose also that $f_j \in L^2(X_j, \mu_j)$ satisfies (65.1) and (65.2) for each j , so that the f_j 's are orthonormal in $L^2(X, \mu)$. If a_1, a_2, \dots is a sequence of real or complex numbers such that $\sum_{j=1}^{\infty} |a_j|^2$ converges, then $\sum_{j=1}^{\infty} a_j f_j$ converges in $L^2(X, \mu)$. If $k \in \mathbf{Z}_+$, $p = 2^k$, and $f_j \in L^p(X_j, \mu_j)$ for each j , with uniformly bounded L^p norm, then it follows from the previous estimates that $\sum_{j=1}^{\infty} a_j f_j$ converges in $L^p(X, \mu)$. More precisely, $\sum_{j=1}^{\infty} a_j f_j$ converges in $L^p(X, \mu)$ in the generalized sense, as in Section 14. In particular, if $f_j \in L^p(X_j, \mu_j)$ for every $j \geq 1$ and $p \in (2, \infty)$, with $\|f_j\|_{L^p(X_j, \mu_j)}$ uniformly bounded in j for each $p > 2$, then $\sum_{j=1}^{\infty} a_j f_j$ converges in $L^p(X, \mu)$ in the generalized sense for each $p \in (2, \infty)$. If the (X_j, μ_j) 's are copies of the same space, and the f_j are copies of the same function on this space, then of course the f_j 's have the same L^p norm for each j .

66 Linear functions on \mathbf{R}^n

Let μ be a Borel probability measure on \mathbf{R}^n that is not the Dirac mass at 0, so that

$$(66.1) \quad \mu(\mathbf{R}^n \setminus \{0\}) > 0.$$

Remember that a linear transformation T from \mathbf{R}^n onto itself is said to be an *orthogonal transformation* if T preserves the standard inner product on \mathbf{R}^n , and hence the standard Euclidean norm on \mathbf{R}^n . Suppose that μ is invariant under orthogonal transformations, in the sense that

$$(66.2) \quad \mu(T(E)) = \mu(E)$$

for every Borel set $E \subseteq \mathbf{R}^n$ and every orthogonal transformation T on \mathbf{R}^n . For example, μ might be surface measure on the unit sphere normalized to have total measure 1, or μ could be absolutely continuous with respect to Lebesgue measure, with a radial density. Also let p be a positive real number, and suppose that

$$(66.3) \quad \int_{\mathbf{R}^n} |x|^p d\mu(x) < \infty.$$

Note that this integral is positive, by hypothesis. If μ is normalized surface measure on the unit sphere, then this condition holds for every $p > 0$. If μ is given by a radial density times Lebesgue measure, then this condition depends on the integrability properties of the density.

Consider

$$(66.4) \quad \lambda_v(x) = \sum_{j=1}^n x_j v_j$$

for each $v \in \mathbf{R}^n$. This is a linear function on \mathbf{R}^n , and every real-valued linear function on \mathbf{R}^n is of this form. By hypothesis, $\lambda_v \in L^p(\mathbf{R}^n, \mu)$ for each $v \in \mathbf{R}^n$. Because of invariance under orthogonal transformations,

$$(66.5) \quad \|\lambda_v\|_{L^p(\mathbf{R}^n, \mu)} = \left(\int_{\mathbf{R}^n} |\lambda_v(x)|^p d\mu(x) \right)^{1/p} = C(p, \mu) |v|,$$

where

$$(66.6) \quad C(p, \mu) = \left(\int_{\mathbf{R}^n} |x_1|^p d\mu(x) \right)^{1/p}$$

and

$$(66.7) \quad |v| = \left(\sum_{j=1}^n v_j^2 \right)^{1/2}$$

is the standard norm on \mathbf{R}^n . Note that $0 < C(p, \mu) < \infty$.

Remember that

$$(66.8) \quad \int_{-\infty}^{\infty} \exp(-t^2) dt = \sqrt{\pi}.$$

To see this, one can begin with

$$\begin{aligned}
 (66.9) \quad \left(\int_{-\infty}^{\infty} \exp(-t^2) dt \right)^2 &= \left(\int_{-\infty}^{\infty} \exp(-t^2) dt \right) \left(\int_{-\infty}^{\infty} \exp(-u^2) du \right) \\
 &= \int_{\mathbf{R}^2} \exp(-t^2 - u^2) dt du.
 \end{aligned}$$

Using polar coordinates, we get that

$$(66.10) \quad \left(\int_{-\infty}^{\infty} \exp(-t^2) dt \right)^2 = 2\pi \int_0^{\infty} r \exp(-r^2) dr.$$

The derivative of $\exp(-r^2)$ is $-2r \exp(-r^2)$, and so

$$(66.11) \quad \int_0^{\infty} 2r \exp(-r^2) dr = 1.$$

This implies (66.8), as desired.

Let μ_n be the measure on \mathbf{R}^n given by $\pi^{-n/2} \exp(-|x|^2)$ times Lebesgue measure. Thus $\mu_n(\mathbf{R}^n) = 1$, by the previous computations, and μ_n is clearly invariant under orthogonal transformations. Also, $|x|^p \in L^p(\mathbf{R}^n, \mu_n)$ for every $p > 0$. Moreover, μ_n is the same as the product of n copies of μ_1 on n copies of \mathbf{R} , as in the previous section.

67 Countability conditions

Remember that a collection β of open subsets of a topological space X is said to be a *base* for the topology of X if for every open set U in X and every point $p \in U$ there is an open set $V \in \beta$ such that $p \in V$ and $V \subseteq U$. In this case,

$$(67.1) \quad U = \bigcup \{V : V \in \beta, V \subseteq U\}$$

for every open set U in X . Conversely, β is a base for the topology of X if every open set in X can be expressed as a union of elements of β . It is especially nice to have a base β for the topology of X with only finitely or countably many elements. This implies that there is a dense set in X with only finitely or countably many elements, by picking an element in each nonempty open set in the base. Conversely, if the topology on X is determined by a metric, and if there is a dense set in X with only finitely or countably many elements, then there is a base for the topology of X with only finitely or countably many elements. More precisely, the collection of open balls in X with centers contained in a dense subset of X and radii of the form $1/n$, $n \in \mathbf{Z}_+$, is a base for the topology of X .

Suppose that β is a base for the topology of X with only finitely or countably many elements, and let $\{U_i\}_{i \in I}$ be a collection of open subsets of X . For each $i \in I$, let β_i be the set of $V \in \beta$ such that $V \subseteq U_i$. Thus

$$(67.2) \quad U_i = \bigcup \{V : V \in \beta_i\}$$

for each $i \in I$, because β is a base for the topology of X . If $\beta' = \bigcup_{i \in I} \beta_i$, then it follows that

$$(67.3) \quad \bigcup_{i \in I} U_i = \bigcup \{V : V \in \beta'\}.$$

For each $V \in \beta'$, let $i(V)$ be an element of I such that $V \subseteq U_{i(V)}$. Also let I' be the set of $i(V)$, $V \in \beta'$. Note that I' has only finitely or countably many elements, because $\beta' \subseteq \beta$ has only finitely or countably many elements. In addition,

$$(67.4) \quad \bigcup_{i \in I'} U_i \subseteq \bigcup_{i \in I} U_i = \bigcup \{V : V \in \beta'\} \subseteq \bigcup U_{i(V)} : V \in \beta' = \bigcup_{i \in I'} U_i,$$

which implies that $\bigcup_{i \in I'} U_i = \bigcup_{i \in I} U_i$.

A set $E \subseteq X$ is said to be σ -compact if there is a sequence K_1, K_2, \dots of compact subsets of X such that $E = \bigcup_{n=1}^{\infty} K_n$. Suppose that X is a locally compact Hausdorff space, and that U is an open set in X . For each $p \in U$, let $U(p)$ be an open set in X such that $p \in U(p)$, $\overline{U(p)}$ is compact, and $\overline{U(p)} \subseteq U$. If there is a base for the topology of X with only finitely or countably many elements, then it follows that there is a set $A \subseteq U$ with only finitely or countably many elements such that $U = \bigcup_{p \in A} U(p)$. Hence $U = \bigcup_{p \in A} \overline{U(p)}$, so that U is σ -compact.

Suppose that X is a locally compact Hausdorff space in which every open set is σ -compact. As in Theorem 2.18 in [130], every positive Borel measure μ on X such that $\mu(K) < \infty$ when $K \subseteq X$ is compact automatically satisfies strong regularity properties. It is easy to see that the real line has this property, for instance, as well as \mathbf{R}^n for every positive integer n . If X is a locally compact Hausdorff space, and there is a base for the topology of X with only finitely or countably many elements, then X has this property, by the remarks in the previous paragraph.

68 Separation conditions

Remember that a topological space X satisfies the *first separation condition* if for every pair of distinct elements p, q of X there is an open set $U \subseteq X$ such that $p \in U$ and $q \notin U$. This is equivalent to asking that every set $A \subseteq X$ with exactly one element be closed, which implies that finite subsets of X are closed. Similarly, X satisfies the *second separation condition* if for every pair p, q of distinct elements of X there are disjoint open subsets U, V of X such that $p \in U, q \in V$. In this case, X is said to be a *Hausdorff* topological space, and X clearly satisfies the first separation condition. If X satisfies the first separation condition and for every point $p \in X$ and closed set $B \subseteq X$ with $p \notin B$ there are disjoint open subsets U, V of X such that $p \in U$ and $B \subseteq V$, then E satisfies the *third separation condition*, and is also said to be *regular*. Note that regular topological spaces are Hausdorff, since one can take $B = \{q\}$ when $q \in X$ and $q \neq p$. If X satisfies the first separation condition and for every pair A, B of

disjoint closed subsets of X there are disjoint open sets U, V such that $A \subseteq U$, $B \subseteq V$, then X satisfies the *fourth separation condition*, and is also said to be *normal*. As before, normal spaces are automatically Hausdorff and regular. It is well known that metric spaces are normal.

Equivalently, X is Hausdorff if for every pair of distinct elements p, q of X there is an open set $U \subseteq X$ such that $p \in U$ and q is not in the closure \overline{U} of U . Similarly, X satisfies the third separation condition if and only if it satisfies the first separation condition and for every point $p \in X$ and open set $W \subseteq X$ with $p \in W$ there is an open set $U \subseteq X$ such that $p \in U$ and $\overline{U} \subseteq W$. This formulation of regularity makes it clear that it is a local property. In the same way, X is normal if and only if for every closed set $A \subseteq X$ and open set $W \subseteq X$ with $A \subseteq W$ there is an open set $U \subseteq X$ such that $A \subseteq U$ and $\overline{U} \subseteq W$.

If X is Hausdorff, then compact subsets of X are closed, and one can show that X satisfies the analogues of regularity and normality for compact sets instead of closed sets. This implies that compact Hausdorff spaces are normal, because closed sets of compact spaces are compact. If X is regular, then one can show that X satisfies the analogue of normality in which at least one of the closed sets is compact. One can also show that locally compact Hausdorff spaces are regular.

It is easy to see that the Cartesian product of a family of topological spaces that satisfy the first or second separation condition has the same property with respect to the product topology. This is because a pair of distinct elements of the product are different in at least one coordinate, and the appropriate separation condition can then be applied in the corresponding space. One can also check that a product of regular spaces is regular. This uses the local characterization of regularity mentioned before.

69 Metrizable

Let $(X, d(x, y))$ be a metric space, and put

$$(69.1) \quad B(p, r) = \{x \in X : d(p, x) < r\}$$

for each $p \in X$ and $r > 0$. This is the open ball in X with center p and radius r , which is well known to be an open set in X , by the triangle inequality. If $A \subseteq X$ and $r > 0$, then

$$(69.2) \quad A_r = \bigcup_{p \in A} B(p, r) = \{x \in X : d(x, p) < r \text{ for some } p \in A\}$$

is an open set in X that contains A . It is easy to check that

$$(69.3) \quad \overline{A} = \bigcap_{r>0} A_r = \bigcap_{n=1}^{\infty} A_{1/n},$$

where \overline{A} denotes the closure of A in X . In particular, every closed set in X can be expressed as the intersection of a sequence of open sets. This implies

that every open set in X can be expressed as the union of a sequence of closed sets. If X is compact, then every closed set in X is compact, and hence every open set in X is σ -compact. If X is σ -compact, then every closed set in X is σ -compact, and it follows that every open set in X is σ -compact as well.

Now let $(X_1, d_1), (X_2, d_2), \dots$ be a sequence of metric spaces, and let $X = \prod_{j=1}^{\infty} X_j$ be their Cartesian product, with the product topology. One can check that

$$(69.4) \quad d(x, y) = \max_{j \geq 1} (\min(d_j(x_j, y_j), 1/j))$$

defines a metric on X for which the corresponding topology is the product topology, where $x = \{x_j\}_{j=1}^{\infty}$, $y = \{y_j\}_{j=1}^{\infty}$. In particular, X may be considered as a compact metric space when X_j is compact for each j .

Uhrhryson's famous metrization theorem implies that there is a metric on a topological space X that determines the same topology when X is regular and there is a countable base for the topology of X . If X is compact, and the topology on X is determined by a metric, then it is easy to show that there is a dense set in X with only finitely or countably many elements, which implies that there is a base for the topology of X with only finitely or countably many elements. This also works when X is σ -compact. Thus a base for the topology of X with only finitely or countably many elements is necessary for metrizability of a compact or σ -compact topological space.

70 Partitions of unity

Let X be a compact Hausdorff topological space. Suppose that for each $p \in X$, we have an open set $U(p)$ in X such that $p \in U(p)$. By Uhrhryson's lemma, there is a nonnegative continuous real-valued function $\phi_p(x)$ on X such that $\phi(p) > 0$ and the support of ϕ_p is contained in $U(p)$. If

$$(70.1) \quad U_1(p) = \{x \in X : \phi_p(x) > 0\},$$

then $U_1(p)$ is an open set in X such that $p \in U_1(p)$ and $U_1(p) \subseteq U(p)$. By compactness, there are finitely many elements p_1, \dots, p_n of X such that

$$(70.2) \quad X = \bigcup_{j=1}^n U_1(p_j).$$

This implies that $\sum_{j=1}^n \phi_{p_j}(x) > 0$ for every $x \in X$. Hence

$$(70.3) \quad \psi_j(x) = \frac{\phi_{p_j}(x)}{\sum_{l=1}^n \phi_{p_l}(x)}$$

defines a nonnegative continuous real-valued function on X . Also,

$$(70.4) \quad \sum_{j=1}^n \psi_j(x) = 1$$

for every $x \in X$, and $\psi_j(x) > 0$ if and only if $\phi_{p_j}(x) > 0$.

As an application, let V be a real or complex vector space equipped with a norm $\|v\|$, and let f be a continuous mapping from X into V . Let $\epsilon > 0$ be given, and let $U(p)$ be an open set in X such that $p \in U(p)$ and

$$(70.5) \quad \|f(x) - f(p)\| < \epsilon$$

for every $x \in U(p)$. Put

$$(70.6) \quad g(x) = \sum_{j=1}^n \psi_j(x) f(p_j),$$

where p_1, \dots, p_n and ψ_1, \dots, ψ_n are as in the previous paragraph. Thus

$$(70.7) \quad \|f(x) - g(x)\| \leq \sum_{j=1}^n \psi_j(x) \|f(x) - f(p_j)\| < \epsilon$$

for every $x \in X$, using (70.5) and the fact that $x \in U(p_j)$ when $\psi_j(x) > 0$. The same argument works when the topology on V is determined by a collection \mathcal{N} of seminorms, and $\|v\|$ is replaced by the maximum of finitely many seminorms in \mathcal{N} .

71 Product spaces

Let X, Y be compact Hausdorff topological spaces, and let $X \times Y$ be their Cartesian product, equipped with the product topology. Thus $X \times Y$ is also a compact Hausdorff space. Also let $f(x, y)$ be a continuous real or complex-valued function on $X \times Y$, and let $\epsilon > 0$ be given. For each $x \in X$ and $y \in Y$, there are open sets $U(x, y) \subseteq X$, $V(x, y) \subseteq Y$ such that $x \in U(x, y)$, $y \in V(x, y)$, and

$$(71.1) \quad |f(u, v) - f(w, z)| < \epsilon$$

for every $u, w \in U(x, y)$ and $v, z \in V(x, y)$, by the continuity of f at (x, y) and the definition of the product topology. If we fix $x \in X$ for a moment, and apply this to each $y \in Y$, then the open sets $V(x, y)$, $y \in Y$, form an open covering of Y . By compactness of Y , there are finitely many elements y_1, \dots, y_n of Y such that

$$(71.2) \quad Y = \bigcup_{j=1}^n V(x, y_j).$$

Put $U(x) = \bigcap_{j=1}^n U(x, y_j)$, so that $U(x)$ is an open set in X that contains x . Moreover,

$$(71.3) \quad |f(u, y) - f(w, y)| < \epsilon$$

for every $u, w \in U(x)$ and $y \in Y$, by applying (71.1) to $v = z = y$, which is contained in $V(y_j)$ for some j . Similarly, one can use compactness of X to show that for every $y \in Y$ there is an open set $V(y) \subseteq X$ such that $y \in V(y)$ and

$$(71.4) \quad |f(x, v) - f(x, z)| < \epsilon$$

for every $v, z \in V(y)$ and $x \in X$.

Let μ, ν be regular Borel probability measures on X, Y , respectively. By the Riesz representation theorem, this is equivalent to having positive linear functionals on the spaces of continuous functions on X, Y that take the value 1 on the constant functions identically equal to 1 on these spaces. If $f(x, y)$ is a continuous function on $X \times Y$, then it follows from the uniform continuity properties in the previous paragraph that

$$(71.5) \quad \int_X f(x, y) d\mu(x), \quad \int_Y f(x, y) d\nu(y)$$

are continuous functions on Y, X , respectively. Thus

$$(71.6) \quad \int_Y \left(\int_X f(x, y) d\mu(x) \right) d\nu(y), \quad \int_X \left(\int_Y f(x, y) d\nu(y) \right) d\mu(x)$$

define nonnegative linear functionals on the space of continuous functions on $X \times Y$ that take the value 1 on the constant function 1. One can also show that these two linear functionals are the same, because they are the same when f is a linear combination of products of continuous functions on X and Y , and because these functions are dense in the space of all continuous functions on $X \times Y$ with respect to the supremum norm. The latter statement can be verified using partitions of unity on X and uniform continuity over Y , for instance, as in the preceding section and paragraph. The Riesz representation theorem implies that there is a unique regular Borel probability measure $\mu \times \nu$ on $X \times Y$ such that this linear functional on the space of continuous functions on $X \times Y$ is given by

$$(71.7) \quad \int_{X \times Y} f(x, y) d(\mu \times \nu)(x, y).$$

There are analogous arguments for nonnegative Borel measures with suitable regularity properties on locally compact Hausdorff spaces, which correspond to nonnegative linear functionals on continuous functions with compact support on these spaces. If the measures are finite, then one can simply compactify the spaces using one-point compactifications.

Let β_X, β_Y be bases for the topologies of X, Y , respectively. It is easy to see that

$$(71.8) \quad \beta_{X \times Y} = \{U \times V : U \in \beta_X, V \in \beta_Y\}$$

is a base for the topology of $X \times Y$. In particular, $\beta_{X \times Y}$ has only finitely or countably many elements when β_X, β_Y have only finitely or countably many elements. In this case, it follows that every open set in $X \times Y$ is the union of finitely or countably many products of open subsets of X and Y . Otherwise, one can check that an open set in $X \times Y$ that is also σ -compact is the union of finitely or countably many products of open subsets of X and Y .

72 Product spaces, 2

Let I be a nonempty set, and suppose that for each $i \in I$ we have a topological space X_i . In practice, we shall be interested in sets I with only finitely or countably many elements. Let $X = \prod_{i \in I} X_i$ be the corresponding Cartesian product, equipped with the product topology.

Suppose that β_i is a base for the topology of X_i for each $i \in I$, and let β be the collection of subsets of X of the form $\prod_{i \in I} U_i$, where $U_i \in \beta_i$ for each $i \in I$, and $U_i = X_i$ for all but finitely many i . It is easy to check that β is a base for the product topology on X . If I has only finitely or countably many elements, and each β_i has only finitely or countably many elements, then β has only finitely or countably many elements too. This follows from the fact that the Cartesian product of finitely many countable sets is countable when I has only finitely many elements. If I is a countably infinite set, then one can use the same argument for finite subsets of I , and apply this to an increasing sequence of finite subsets of I whose union is all of I .

If X_i is Hausdorff for each $i \in I$, then X is Hausdorff. If X_i is compact for each $i \in I$, then X is compact, by Tychonoff's theorem. Of course, this is much more elementary when I has only finitely many elements. If I has only finitely or countably many elements and each X_i is metrizable, then X is metrizable, and compactness can be handled in a simpler way using sequential compactness. This approach can also be applied directly when I has only finitely or countably many elements and there is a base for the topology of X_i with only finitely or countably many elements for each $i \in I$, so that there is also a base for the topology of X with only finitely or countably many elements.

Let f be a continuous real or complex-valued function on X . For each $\epsilon > 0$ and $x \in X$, there is an open set $U(x)$ in X such that $x \in U(x)$ and

$$(72.1) \quad |f(y) - f(z)| < \epsilon$$

for every $y, z \in U(x)$. More precisely, we can take $U(x)$ to be a basic open set in the product topology, so that there is a finite set $I(x) \subseteq I$ such that $U(x) = \prod_{i \in I} U_i(x)$ for some open sets $U_i(x) \subseteq X_i$, where $U_i(x) = X_i$ for every $i \in I \setminus I(x)$. In particular, if $y \in U(x)$, $z \in X$, and $y_i = z_i$ for each $i \in I(x)$, then it follows that $z \in U(x)$, and hence (72.1) holds.

If X_i is compact for each $i \in I$, so that X is compact, then there are finitely many elements $x(1), \dots, x(n)$ of X such that

$$(72.2) \quad X = \bigcup_{j=1}^n U(x(j)).$$

Put $I_\epsilon = \bigcup_{j=1}^n I(x(j))$, so that $I_\epsilon \subseteq I$ has only finitely many elements. If $y, z \in X$ satisfy $y_i = z_i$ for every $i \in I_\epsilon$, then it is easy to see that (72.1) holds. This is because $y \in U(x(j))$ for some $j = 1, \dots, n$, and so $z \in U(x(j))$ too. Thus continuous functions on X may be approximated uniformly by functions of finitely many variables under these conditions.

Suppose that μ_i is a regular Borel probability measure on X_i for each i . If $A \subseteq I$ is a nonempty set with only finitely many elements, then let $L_A(f)$ be the function on X which is constant in x_i for each $i \in A$ obtained by integrating f in x_i with respect to μ_i for each $i \in A$. If $A \cap I_\epsilon = \emptyset$, then

$$(72.3) \quad |L_A(f)(y) - f(y)| < \epsilon$$

for every $y \in X$, since (72.1) holds for every $z \in X$ such that $y_i = z_i$ when $i \in I \setminus A$. If $A, B \subseteq I$ are finite sets such that $I_\epsilon \subseteq A, B$, then

$$(72.4) \quad |L_A(f)(y) - L_B(f)(y)| < 2\epsilon$$

for every $y \in X$. This uses the previous estimate applied to $A \setminus B$ and $B \setminus A$, to estimate the difference between each of $L_A(f)$, $L_B(f)$ and $L_{A \cap B}(f)$.

Let \mathcal{A} be the collection of all finite subsets of I , ordered by inclusion. This is a directed system, because for every $A, B \in \mathcal{A}$ we have that $A \cup B \in \mathcal{A}$ and $A, B \subseteq A \cup B$. If f is a continuous function on X , then one can think of $\{L_A(f)\}_{A \in \mathcal{A}}$ as a net of functions on X indexed by \mathcal{A} . One can show that this net converges uniformly to a constant on X for every continuous function on X . This uses the fact that the net satisfies a uniform Cauchy condition on X , as in the previous paragraph.

In the limit, we get a positive linear functional on the space of continuous functions on X which takes the value 1 on the constant function 1. The Riesz representation theorem implies that this linear functional can be expressed in terms of a unique regular Borel probability measure on X , which corresponds to the product of the μ_i 's. As usual, the situation is especially nice when I is countably infinite, and each X_i has a base β_i for its topology with only finitely or countably many elements. This leads to a base β for the topology of X consisting of only finitely or countably many basic open sets in X , as before, which implies in particular that every open set in X is the union of finitely or countably many basic open sets. Otherwise, every open set in X that is also σ -compact is the union of finitely or countably many basic open sets, as in the previous section.

Part III

Conditional expectation and martingales

73 σ -Subalgebras

Let (X, \mathcal{A}, μ) be a probability space, and let \mathcal{B} be a σ -subalgebra of \mathcal{A} . Thus (X, \mathcal{B}, μ) is also a probability space, where the measure μ is restricted to \mathcal{B} . If a real or complex-valued function f on X is measurable with respect to \mathcal{B} , then

it is automatically measurable with respect to \mathcal{A} as well. If f is measurable with respect to \mathcal{B} and integrable with respect to μ , then f is also integrable as a function which is measurable with respect to \mathcal{A} , and the integral

$$(73.1) \quad \int_X f d\mu$$

is the same with respect to both \mathcal{A} and \mathcal{B} .

For example, \mathcal{B} might consist of only the empty set \emptyset and X itself, in which case the only functions on X that are measurable with respect to \mathcal{B} are constant functions. As another example, one might take X to be the closed unit interval $[0, 1]$, \mathcal{A} to be the σ -algebra of Lebesgue measurable subsets of $[0, 1]$, μ to be Lebesgue measure on $[0, 1]$, and \mathcal{B} to be the σ -algebra of Borel subsets of $[0, 1]$. It is well known that for each Lebesgue measurable set $A \subseteq [0, 1]$ there are Borel sets $B_1, B_2 \subseteq [0, 1]$ such that $B_1 \subseteq A \subseteq B_2$ and $\mu(B_2 \setminus B_1) = 0$. More precisely, one can take B_1 to be a countable union of compact sets, and B_2 to be a countable intersection of relatively open sets in $[0, 1]$.

Let $(X_1, \mathcal{A}_1, \mu_1)$, $(X_2, \mathcal{A}_2, \mu_2)$ be probability spaces, and let $X = X_1 \times X_2$ be their Cartesian product, with the corresponding product measure $\mu_1 \times \mu_2$ and σ -algebra \mathcal{A} . Let \mathcal{B}_1 be the collection of subsets of X of the form $E \times X_2$ with $E \in \mathcal{A}_1$, and let \mathcal{B}_2 be the collection of subsets of X of the form $X_1 \times E$ with $E \in \mathcal{A}_2$. It is easy to see that $\mathcal{B}_1, \mathcal{B}_2$ are σ -subalgebras of \mathcal{A} , and that a function $f(x_1, x_2)$ on X is measurable with respect to \mathcal{B}_1 or \mathcal{B}_2 if and only if it is measurable with respect to \mathcal{A} and constant in x_2 or x_1 , respectively. Thus measurable functions on X with respect to $\mathcal{B}_1, \mathcal{B}_2$ may be identified with functions on X_1, X_2 that are measurable with respect to $\mathcal{A}_1, \mathcal{A}_2$, respectively.

As a variant of this, suppose that X_1, X_2 are topological spaces, and let $X = X_1 \times X_2$ be equipped with the product topology. If $A_1 \subseteq X_1, A_2 \subseteq X_2$ are Borel sets, then $A_1 \times X_2, X_1 \times A_2$ are Borel sets in X , by standard reasoning. In particular,

$$(73.2) \quad A_1 \times A_2 = (A_1 \times X_2) \cap (X_1 \times A_2)$$

is a Borel set in X . At any rate, the collections of subsets of X of the form $A_1 \times X_2, X_1 \times A_2$, where A_1, A_2 are Borel subsets of X_1, X_2 , respectively, are σ -subalgebras of the Borel sets in X . As in Section 71, if there are bases for the topologies of X_1, X_2 with only finitely or countably many elements, then every open set in X is the union of finitely or countably many products of open subsets of X_1 and X_2 . This implies that every open set in X is in the σ -algebra generated by products of Borel sets in X_1, X_2 , and hence that every Borel set in X is in this σ -algebra. It follows that the σ -algebra of subsets of X generated by products of Borel sets in X_1, X_2 is the same as the σ -algebra of Borel sets in X under these conditions.

74 L^p Spaces

Let (X, \mathcal{A}, μ) be a probability space, and let \mathcal{B} be a σ -subalgebra of \mathcal{A} . If f, g are measurable functions on X with respect to \mathcal{A} , then

$$(74.1) \quad \{x \in X : f(x) = g(x)\}$$

is a measurable set in X with respect to \mathcal{A} . If f, g are measurable with respect to \mathcal{B} , then (74.1) is measurable with respect to \mathcal{B} . Of course, f and g are said to be equal almost everywhere with respect to μ when

$$(74.2) \quad \mu(\{x \in X : f(x) \neq g(x)\}) = 0.$$

Let $L^p(X, \mathcal{A}), L^p(X, \mathcal{B})$ be the L^p spaces of measurable functions on X with respect to \mathcal{A}, \mathcal{B} , for $0 < p \leq \infty$. These spaces also involve the measure μ , but we omit this from the notation when it is unambiguous. Because measurable functions on X with respect to \mathcal{B} are also measurable with respect to \mathcal{A} , we get an isometric linear embedding of $L^p(X, \mathcal{B})$ into $L^p(X, \mathcal{A})$ for each $p, 0 < p \leq \infty$.

Note that $L^p(X, \mathcal{B})$ corresponds to a closed linear subspace of $L^p(X, \mathcal{A})$ for each $p, 0 < p \leq \infty$. One way to see this is to use the completeness of $L^p(X, \mathcal{B})$ and the fact that the embedding into $L^p(X, \mathcal{A})$ is isometric. Basically the same argument can be given more explicitly as follows. Suppose that $\{f_j\}_{j=1}^\infty$ is a sequence of elements of $L^p(X, \mathcal{B})$ that converges in the L^p norm to $f \in L^p(X, \mathcal{A})$. By passing to a subsequence, we may suppose that $\{f_j\}_{j=1}^\infty$ converges pointwise almost everywhere to f . It is well known that the set of $x \in X$ such that $\{f_j(x)\}_{j=1}^\infty$ converges in \mathbf{R} or \mathbf{C} , as appropriate, is measurable with respect to \mathcal{B} , because each f_j is measurable with respect to \mathcal{B} . The complement of this set has measure 0 by hypothesis, and we may suppose that $\{f_j(x)\}_{j=1}^\infty$ converges in \mathbf{R} or \mathbf{C} for every $x \in X$, by setting $f_j(x) = 0$ on the set where the sequence does not converge initially. The limit is automatically measurable with respect to \mathcal{B} , and equal to f almost everywhere. This shows that f is in the image of $L^p(X, \mathcal{B})$ in $L^p(X, \mathcal{A})$, as desired.

75 Conditional expectation

Let (X, \mathcal{A}, μ) be a probability space, and let \mathcal{B} be a σ -subalgebra of \mathcal{A} . If $f \in L^1(X, \mathcal{A})$, then

$$(75.1) \quad \mu_f(A) = \int_A f d\mu$$

defines a real or complex measure on \mathcal{A} , as appropriate. By construction, μ_f is absolutely continuous with respect to μ . Hence the restriction of μ_f to \mathcal{B} is absolutely continuous with respect to the restriction of μ to \mathcal{B} . The Radon–Nikodym theorem implies that there is a measurable function $f_{\mathcal{B}}$ on X with respect to \mathcal{B} which is integrable with respect to μ and satisfies

$$(75.2) \quad \mu_f(B) = \int_B f_{\mathcal{B}} d\mu$$

for every $B \in \mathcal{B}$. If f'_B is another measurable function on X with respect to \mathcal{B} which is integrable with respect to μ and satisfies

$$(75.3) \quad \mu_f(B) = \int_B f'_B d\mu$$

for every $B \in \mathcal{B}$, then it is easy to see that $f'_B = f_B$ almost everywhere with respect to μ . Thus f_B is uniquely determined as an element of $L^1(X, \mathcal{B})$. This function f_B is known as the *conditional expectation* of f with respect to \mathcal{B} , and may be denoted $E(f | \mathcal{B})$.

For example, if $\mathcal{B} = \{\emptyset, X\}$, so that only constant functions are measurable with respect to \mathcal{B} , then $E(f | \mathcal{B})$ reduces to the ordinary expectation

$$(75.4) \quad E(f) = \int_X f d\mu.$$

If $\mathcal{A} = \mathcal{B}$, then $f_B = f$. For any \mathcal{A}, \mathcal{B} , we can take $f_B = f$ when f is measurable with respect to \mathcal{B} .

Let $(X_1, \mathcal{A}_1, \mu_1), (X_2, \mathcal{A}_2, \mu_2)$ be probability spaces, and let $X = X_1 \times X_2$ with the product measure $\mu = \mu_1 \times \mu_2$ and corresponding σ -algebra \mathcal{A} . Also let $\mathcal{B}_1, \mathcal{B}_2$ be the σ -subalgebras of \mathcal{A} defined in Section 73. If $f(x_1, x_2) \in L^1(X, \mathcal{A})$, then

$$(75.5) \quad f_1(x_1) = \int_{X_2} f(x_1, x_2) d\mu_2(x_2),$$

$$(75.6) \quad f_2(x_2) = \int_{X_1} f(x_1, x_2) d\mu_1(x_1)$$

are defined almost everywhere on X_1, X_2 , respectively, and determine integrable functions on these spaces, as in Fubini's theorem. In this case,

$$(75.7) \quad f_{\mathcal{B}_1}(x_1, x_2) = f_1(x_1), \quad f_{\mathcal{B}_2}(x_1, x_2) = f_2(x_2)$$

are measurable functions on X with respect to $\mathcal{B}_1, \mathcal{B}_2$, respectively, and satisfy the requirements of the conditional expectation, again by Fubini's theorem.

76 Product spaces, 3

Let X_1, X_2 be compact Hausdorff topological spaces, and let $X = X_1 \times X_2$ be their Cartesian product, with the product topology. Also let μ_1, μ_2 be regular Borel probability measures on X_1, X_2 , respectively, which may be given by positive linear functionals on the spaces of continuous functions on X_1, X_2 that take the value 1 on the constant functions equal to 1 on these spaces, by the Riesz representation theorem. If $f(x_1, x_2)$ is a continuous function on X , then

$$(76.1) \quad f_1(x_1) = \int_{X_2} f(x_1, x_2) d\mu_2(x_2), \quad f_2(x_2) = \int_{X_1} f(x_1, x_2) d\mu_1(x_1)$$

are continuous functions on X_1 , X_2 , respectively, by the uniform continuity properties of $f(x_1, x_2)$ in each variable separately discussed in Section 71. In addition,

$$(76.2) \quad \int_{X_1} f_1(x_1) d\mu_1(x_1) = \int_{X_2} f_2(x_2) d\mu_2(x_2)$$

defines a positive linear functional on the space of continuous functions on X that takes the value 1 on the constant 1, and hence determines a regular Borel probability measure μ on X by the Riesz representation theorem, as in Section 71 again.

In this context, one can think of μ_f as the regular Borel measure on X determined by

$$(76.3) \quad \phi \mapsto \int_X \phi f d\mu,$$

as a bounded linear functional on the space of continuous functions on X . If ψ is a continuous function on X_1 , which can also be considered as a continuous function on X that is constant in x_2 , then this linear functional applied to $\phi(x_1, x_2) = \psi(x_1)$ reduces to

$$(76.4) \quad \int_{X_1} \psi f_1 d\mu_1 = \int_X \psi f_1 d\mu.$$

Of course, there is an analogous statement for continuous functions on X_2 . In this way, conditional expectation can be expressed more directly in terms of linear functionals on continuous functions.

77 Measurable partitions

Let (X, \mathcal{A}, μ) be a probability space, and let \mathcal{P} be a partition of X consisting of finitely or countably many measurable subsets of X . Thus the elements of \mathcal{P} are pairwise-disjoint measurable subsets of X whose union is all of X . Let $\mathcal{B} = \mathcal{B}(\mathcal{P})$ be the collection of subsets of X that can be expressed as unions of elements of \mathcal{P} , including the empty set. It is easy to see that \mathcal{B} is a σ -subalgebra of \mathcal{A} , and that a function f on X is measurable with respect to \mathcal{P} if and only if f is constant on each of the elements of \mathcal{P} .

If $f \in L^1(X, \mathcal{A})$, then one can check that

$$(77.1) \quad f_{\mathcal{B}}(x) = \frac{1}{\mu(A)} \int_A f d\mu$$

when $x \in A \in \mathcal{P}$ and $\mu(A) > 0$. Let us ask that $\mu(A) > 0$ for every $A \in \mathcal{P}$, for the sake of simplicity. Thus $f_{\mathcal{B}}(x)$ is defined for every $x \in X$ by this expression, and is constant on elements of \mathcal{P} , and hence is measurable with respect to \mathcal{B} .

If ν is a real or complex measure on \mathcal{A} , then the restriction of ν to a σ -subalgebra \mathcal{B} of \mathcal{A} may be absolutely continuous with respect to the restriction of μ to \mathcal{B} , even if ν is not absolutely continuous with respect to μ on \mathcal{A} . In this

case, the Radon–Nikodym theorem implies that there is a unique $f_{\mathcal{B}} \in L^1(X, \mathcal{B})$ such that

$$(77.2) \quad \nu(B) = \int_B f_{\mathcal{B}} d\mu$$

for every $B \in \mathcal{B}$, as before. If $\mathcal{B} = \mathcal{B}(\mathcal{P})$ and $\mu(A) > 0$ for every $A \in \mathcal{P}$, then any measure on \mathcal{B} is absolutely continuous with respect to the restriction of μ to \mathcal{B} . As in the previous situation,

$$(77.3) \quad f_{\mathcal{B}}(x) = \frac{\nu(A)}{\mu(A)}$$

for every $x \in A \in \mathcal{P}$.

78 Basic properties

Let (X, \mathcal{B}, μ) be a measure space, and let g_0 be a real-valued integrable function on X . If

$$(78.1) \quad \int_B g_0 d\mu \geq 0$$

for every $B \in \mathcal{B}$, then $g_0 \geq 0$ almost everywhere on X . To see this, put

$$(78.2) \quad B_0 = \{x \in X : g_0(x) < 0\}.$$

If $\mu(B_0) > 0$, then it follows that

$$(78.3) \quad \int_{B_0} g_0 d\mu < 0,$$

a contradiction.

Suppose now that g is a real or complex-valued integrable function on X , and that h is a nonnegative real-valued integrable function on X such that

$$(78.4) \quad \left| \int_B g d\mu \right| \leq \int_B h d\mu$$

for every $B \in \mathcal{B}$. We would like to check that $|g| \leq h$ almost everywhere on X under these conditions. If g is real-valued, then we can apply the previous argument to $h \pm g$, to get that $h \pm g \geq 0$ almost everywhere on X . If g is complex-valued, then the same argument shows that $\operatorname{Re} \alpha g \leq h$ almost everywhere on X for every $\alpha \in \mathbf{C}$ with $|\alpha| = 1$. This implies that $|g| \leq h$ almost everywhere, by using a countable dense set of α 's in the unit circle.

Now let (X, \mathcal{A}, μ) be a probability space, and let \mathcal{B} be a σ -subalgebra of \mathcal{A} . If $f \in L^1(X, \mathcal{A})$ is real-valued and nonnegative, then

$$(78.5) \quad \int_B f_{\mathcal{B}} d\mu = \int_B f d\mu \geq 0$$

for every $B \in \mathcal{B}$. This implies that $f_{\mathcal{B}} \geq 0$ almost everywhere on X , by the argument at the beginning of the section. Of course, it is important here that $f_{\mathcal{B}}$ is also measurable with respect to \mathcal{B} . Similarly, if $f > 0$ almost everywhere on X , then

$$(78.6) \quad \int_B f_{\mathcal{B}} d\mu = \int_B f d\mu > 0$$

for every $B \in \mathcal{B}$ with $\mu(B) > 0$, and one can use this to show that $f_{\mathcal{B}} > 0$ almost everywhere on X .

If f is any integrable function on X that is measurable with respect to \mathcal{A} , then we can apply the preceding observation to $|f|$ to get that

$$(78.7) \quad |f|_{\mathcal{B}} = E(|f| \mid \mathcal{B}) \geq 0$$

almost everywhere on X . Moreover,

$$(78.8) \quad \left| \int_B f_{\mathcal{B}} d\mu \right| = \left| \int_B f d\mu \right| \leq \int_B |f| d\mu = \int_B |f|_{\mathcal{B}} d\mu$$

for every $B \in \mathcal{B}$, which implies that

$$(78.9) \quad |f_{\mathcal{B}}| \leq |f|_{\mathcal{B}}$$

almost everywhere on X , by the earlier remarks. As before, it is important here that both $f_{\mathcal{B}}$ and $|f|_{\mathcal{B}}$ are measurable with respect to \mathcal{B} , to apply the arguments at the beginning of the section. In particular,

$$(78.10) \quad \int_X |f_{\mathcal{B}}| d\mu \leq \int_X |f|_{\mathcal{B}} d\mu = \int_X |f| d\mu,$$

using the fact that $X \in \mathcal{B}$ in the last step.

Alternatively, let ν be a real or complex measure on \mathcal{A} , and let $|\nu|$ be the corresponding total variation measure on \mathcal{A} . Also let $\nu_{\mathcal{B}}$ be the restriction of ν to \mathcal{B} , and let $|\nu_{\mathcal{B}}|$ be its total variation, as a measure on \mathcal{B} . It is easy to see that

$$(78.11) \quad |\nu_{\mathcal{B}}|(B) \leq |\nu|(B)$$

for every $B \in \mathcal{B}$, so that $|\nu_{\mathcal{B}}|$ is less than or equal to the restriction of $|\nu|$ to \mathcal{B} . If $f \in L^1(X, \mathcal{A})$ and μ_f is as in (75.1), then one can show that $|\mu_f| = \mu_{|f|}$. This gives another way to look at (78.9), since the restriction of μ_f to \mathcal{B} is given by integrating $f_{\mathcal{B}}$.

Note that $f \mapsto f_{\mathcal{B}}$ defines a linear mapping from $L^1(X, \mathcal{A})$ into $L^1(X, \mathcal{B})$, because of the uniqueness of the conditional expectation. More precisely, this mapping sends $L^1(X, \mathcal{A})$ onto $L^1(X, \mathcal{B})$, because $f_{\mathcal{B}} = f$ when f is measurable with respect to \mathcal{B} . If f, f' are real-valued integrable functions on X that are measurable with respect to \mathcal{A} and satisfy $f \leq f'$ almost everywhere on X , then

$$(78.12) \quad f_{\mathcal{B}} \leq f'_{\mathcal{B}}$$

almost everywhere on X . This follows from the linearity of the conditional expectation and the fact that $f' - f \geq 0$ almost everywhere, so that $(f' - f)_B \geq 0$ almost everywhere on X . If f is a real or complex-valued integrable function on X and f' is a nonnegative real-valued integrable function on X such that $|f| \leq f'$ almost everywhere, then we get that

$$(78.13) \quad |f_B| \leq |f|_B \leq f'_B$$

almost everywhere on X . In particular, this holds when f' is a constant, in which case f'_B is the same constant. This implies that $f_B \in L^\infty(X, \mathcal{B})$ when $f \in L^\infty(X, \mathcal{A})$, with

$$(78.14) \quad \|f_B\|_\infty \leq \|f\|_\infty.$$

Let f be a real-valued integrable function on X that is measurable with respect to \mathcal{A} and takes values in an interval $I \subseteq \mathbf{R}$ almost everywhere. This interval may be open, closed, or half-open and half-closed, and it may also be unbounded, such as a half-line or the whole real line. One can check that f_B takes values in I almost everywhere as well, by comparing f with constant functions. If $\phi : I \rightarrow \mathbf{R}$ is convex and $\phi \circ f$ is integrable on X , then Jensen's inequality implies that

$$(78.15) \quad \phi\left(\frac{1}{\mu(A)} \int_A f \, d\mu\right) \leq \frac{1}{\mu(A)} \int_A \phi \circ f \, d\mu$$

for every $A \in \mathcal{A}$ with $\mu(A) > 0$. Hence

$$(78.16) \quad \phi\left(\frac{1}{\mu(B)} \int_B f_B \, d\mu\right) \leq \frac{1}{\mu(B)} \int_B (\phi \circ f)_B \, d\mu$$

for every $B \in \mathcal{B}$ with $\mu(B) > 0$, because these averages can be reduced to those in (78.15). Using this, one can check that

$$(78.17) \quad \phi(f_B) \leq (\phi \circ f)_B$$

almost everywhere on X . More precisely, one can apply the previous inequality for averages to sets $B \in \mathcal{B}$ where f_B , $(\phi \circ f)_B$ are approximately constant.

Of course, $\phi(t) = |t|^p$ is a convex function on the real line when $1 \leq p < \infty$. If $f \in L^p(X, \mathcal{A})$ is real-valued, then we get that

$$(78.18) \quad |f_B|^p \leq (|f|^p)_B = E(|f|^p \mid \mathcal{B})$$

almost everywhere on X , as in the previous paragraph. If f is complex-valued, then one can apply this to $|f|$, to get that

$$(78.19) \quad |f_B|^p \leq (|f|_B)^p \leq (|f|^p)_B,$$

using (78.9) in the first step. It follows that

$$(78.20) \quad \int_X |f_B|^p \, d\mu \leq \int_X (|f|^p)_B \, d\mu = \int_X |f|^p \, d\mu,$$

because $X \in \mathcal{B}$, and that $f_{\mathcal{B}} \in L^p(X, \mathcal{B})$ in particular. Equivalently,

$$(78.21) \quad \|f_{\mathcal{B}}\|_p \leq \|f\|_p,$$

which also holds when $p = \infty$, as in (78.14).

Remember that $\mathbf{1}_E(x)$ denotes the indicator function of a set $E \subseteq X$, equal to 1 when $x \in E$ and to 0 when $x \in X \setminus E$. If $f \in L^1(X, \mathcal{A})$ and $A, E \in \mathcal{A}$, then of course

$$(78.22) \quad \int_A f \mathbf{1}_E d\mu = \int_{A \cap E} f d\mu.$$

If $B, E \in \mathcal{B}$, then $B \cap E \in \mathcal{B}$, and

$$(78.23) \quad \begin{aligned} \int_B (f \mathbf{1}_E)_{\mathcal{B}} d\mu &= \int_B f \mathbf{1}_E d\mu = \int_{B \cap E} f d\mu \\ &= \int_{B \cap E} f_{\mathcal{B}} d\mu = \int_B f_{\mathcal{B}} \mathbf{1}_E d\mu. \end{aligned}$$

This implies that

$$(78.24) \quad (f \mathbf{1}_E)_{\mathcal{B}} = f_{\mathcal{B}} \mathbf{1}_E,$$

since $f_{\mathcal{B}} \mathbf{1}_E$ is measurable with respect to \mathcal{B} . Similarly, if $g \in L^\infty(X, \mathcal{B})$, then

$$(78.25) \quad (f g)_{\mathcal{B}} = f_{\mathcal{B}} g.$$

This follows from the previous statement by approximating g by simple functions that are measurable with respect to \mathcal{B} . If $f \in L^p(X, \mathcal{A})$, $1 \leq p \leq \infty$, then (78.25) also works for $g \in L^q(X, \mathcal{B})$, where $1/p + 1/q = 1$, by the same argument.

Note that $f_{\mathcal{B}} = 0$ almost everywhere on X if and only if

$$(78.26) \quad \int_B f d\mu = 0$$

for every $B \in \mathcal{B}$. If $f \in L^p(X, \mathcal{A})$, $1 \leq p \leq \infty$, then this implies that

$$(78.27) \quad \int_X f g d\mu = 0$$

for every $g \in L^q(X, \mathcal{B})$, where $1/p + 1/q = 1$ again. This uses the fact that simple functions are dense in $L^q(X, \mathcal{B})$. If $p = 2$, then the collection of $f \in L^2(X, \mathcal{A})$ such that $f_{\mathcal{B}} = 0$ is the same as the orthogonal complement of $L^2(X, \mathcal{B})$ as a linear subspace of $L^2(X, \mathcal{A})$, and $f \mapsto f_{\mathcal{B}}$ is the same as the orthogonal projection of $L^2(X, \mathcal{A})$ onto $L^2(X, \mathcal{B})$.

Suppose now that $\mathcal{B}_1, \mathcal{B}_2$ are σ -subalgebras of \mathcal{A} , with $\mathcal{B}_1 \subseteq \mathcal{B}_2$. If f is an integrable function on X with respect to \mathcal{A} , then

$$(78.28) \quad (f_{\mathcal{B}_2})_{\mathcal{B}_1} = f_{\mathcal{B}_1}.$$

To see this, let $B \in \mathcal{B}_1$ be given, and observe that

$$(78.29) \quad \int_B (f_{\mathcal{B}_2})_{\mathcal{B}_1} d\mu = \int_B f_{\mathcal{B}_2} d\mu = \int_B f d\mu = \int_B f_{\mathcal{B}_1} d\mu,$$

because $B \in \mathcal{B}_2$ as well. This corresponds to the fact that restricting a measure ν on \mathcal{A} to \mathcal{B}_1 is the same as restricting ν to \mathcal{B}_2 , and then to \mathcal{B}_1 .

79 Distances between measurable sets

Remember that the symmetric difference $A \triangle B$ of two sets A, B is defined by

$$(79.1) \quad A \triangle B = (A \setminus B) \cup (B \setminus A).$$

If C is another set, then it is easy to see that

$$(79.2) \quad A \triangle C \subseteq (A \triangle B) \cup (B \triangle C).$$

Let (X, \mathcal{A}, μ) be a probability space, and define $d(A, B)$ for $A, B \in \mathcal{A}$ by

$$(79.3) \quad d(A, B) = \mu(A \triangle B).$$

Thus $d(A, A) = 0$, $d(A, B) = d(B, A) \geq 0$, and

$$(79.4) \quad d(A, C) \leq d(A, B) + d(B, C)$$

for every $A, B, C \in \mathcal{A}$, by (79.2). This shows that $d(A, B)$ is a semimetric on \mathcal{A} , which means that it satisfies all of the requirements of a metric, except that $d(A, B) = 0$ may not imply that $A = B$. In this case, $d(A, B) = 0$ when A and B are the same up to sets of measure 0. Equivalently, $d(A, B)$ is equal to the distance between the indicator functions $\mathbf{1}_A, \mathbf{1}_B$ in L^1 .

Observe that $(X \setminus A) \triangle (X \setminus B) = A \triangle B$ for every $A, B \subseteq X$, and hence

$$(79.5) \quad d(X \setminus A, X \setminus B) = d(A, B)$$

when $A, B \in \mathcal{A}$. Moreover,

$$\begin{aligned} (79.6) \quad & (A_1 \cup A_2) \triangle (B_1 \cup B_2) \\ &= ((A_1 \cup A_2) \setminus (B_1 \cup B_2)) \cup ((B_1 \cup B_2) \setminus (A_1 \cup A_2)) \\ &= (A_1 \setminus (B_1 \cup B_2)) \cup (A_2 \setminus (B_1 \cup B_2)) \cup (B_1 \setminus (A_1 \cup A_2)) \cup (B_2 \setminus (A_1 \cup A_2)) \\ &\subseteq (A_1 \setminus B_1) \cup (A_2 \setminus B_2) \cup (B_1 \setminus A_1) \cup (B_2 \setminus A_2) \\ &= (A_1 \triangle B_1) \cup (A_2 \triangle B_2) \end{aligned}$$

for every $A_1, A_2, B_1, B_2 \subseteq X$. Therefore

$$(79.7) \quad d(A_1 \cup A_2, B_1 \cup B_2) \leq d(A_1, B_1) + d(A_2, B_2)$$

when $A_1, A_2, B_1, B_2 \in \mathcal{A}$. This implies that

$$(79.8) \quad d(A_1 \cap A_2, B_1 \cap B_2) \leq d(A_1, B_1) + d(A_2, B_2)$$

for every $A_1, A_2, B_1, B_2 \in \mathcal{A}$, because

$$(79.9) \quad X \setminus (A_1 \cap A_2) = (X \setminus A_1) \cup (X \setminus A_2),$$

and similarly for $X \setminus (B_1 \cap B_2)$. This also uses (79.5) applied to $A_1 \cap A_2, B_1 \cap B_2$ instead of A, B , and then to A_1, B_1 and A_2, B_2 .

If $A_1 \subseteq A_2 \subseteq \cdots$ is an increasing sequence of measurable subsets of X , then $\{A_j\}_{j=1}^\infty$ converges to their union $\bigcup_{j=1}^\infty A_j$ with respect to $d(A, B)$, in the sense that

$$(79.10) \quad \lim_{n \rightarrow \infty} d\left(A_n, \bigcup_{j=1}^\infty A_j\right) = 0.$$

To see this, note that $A_n \subseteq \bigcup_{j=1}^\infty A_j$ for each n , so that

$$(79.11) \quad A_n \triangle \left(\bigcup_{j=1}^\infty A_j\right) = \left(\bigcup_{j=1}^\infty A_j\right) \setminus A_n = \bigcup_{j=n}^\infty (A_{j+1} \setminus A_j).$$

Hence

$$(79.12) \quad d\left(A_n, \bigcup_{j=1}^\infty A_j\right) = \sum_{j=n}^\infty \mu(A_{j+1} \setminus A_j).$$

Of course, the sets $A_{j+1} \setminus A_j$ are pairwise disjoint, and so $\sum_{j=1}^\infty \mu(A_{j+1} \setminus A_j)$ converges, by countable additivity. This implies that

$$(79.13) \quad \lim_{n \rightarrow \infty} \sum_{j=n}^\infty \mu(A_{j+1} \setminus A_j) = 0,$$

as desired. Similarly, if $B_1 \supseteq B_2 \supseteq \cdots$ is a decreasing sequence of measurable sets, then $\{B_j\}_{j=1}^\infty$ converges to $\bigcap_{j=1}^\infty B_j$ with respect to $d(A, B)$, in the sense that

$$(79.14) \quad \lim_{n \rightarrow \infty} d\left(B_n, \bigcap_{j=1}^\infty B_j\right) = 0.$$

This follows from the previous case applied to $A_j = X \setminus B_j$.

Let $\{A_j\}_{j=1}^\infty$ be a sequence of subsets of X , and put

$$(79.15) \quad B_k = \bigcup_{j=k}^\infty A_j, \quad C_l = \bigcap_{j=l}^\infty A_j$$

for each $k, l \geq 1$. Thus

$$(79.16) \quad B_{k+1} \subseteq B_k, \quad C_l \subseteq C_{l+1}, \quad \text{and} \quad C_k \subseteq B_k$$

for each k, l . The upper and lower limits of $\{A_j\}_{j=1}^\infty$ are the subsets of X defined by

$$(79.17) \quad \limsup_{j \rightarrow \infty} A_j = \bigcap_{k=1}^\infty B_k, \quad \liminf_{j \rightarrow \infty} A_j = \bigcup_{l=1}^\infty C_l.$$

In particular,

$$(79.18) \quad \liminf_{j \rightarrow \infty} A_j \subseteq \limsup_{j \rightarrow \infty} A_j.$$

Suppose that $A_j \in \mathcal{A}$ for each j , so that $B_k, C_l \in \mathcal{A}$ for every k, l , and hence

$$(79.19) \quad \limsup_{j \rightarrow \infty} A_j, \liminf_{j \rightarrow \infty} A_j \in \mathcal{A}.$$

Because of monotonicity,

$$(79.20) \quad \lim_{k \rightarrow \infty} \mu(B_k) = \mu\left(\limsup_{j \rightarrow \infty} A_j\right), \quad \lim_{l \rightarrow \infty} \mu(C_l) = \mu\left(\liminf_{j \rightarrow \infty} A_j\right).$$

It follows that

$$(79.21) \quad \mu\left(\limsup_{j \rightarrow \infty} A_j\right) = \mu\left(\liminf_{j \rightarrow \infty} A_j\right)$$

if and only if

$$(79.22) \quad \lim_{n \rightarrow \infty} \mu(B_n \setminus C_n) = 0.$$

If this condition holds and $A \in \mathcal{A}$ satisfies

$$(79.23) \quad \liminf_{j \rightarrow \infty} A_j \subseteq A \subseteq \limsup_{j \rightarrow \infty} A_j,$$

then it is easy to see that

$$(79.24) \quad \lim_{n \rightarrow \infty} d(A_n, A) = 0.$$

More precisely,

$$(79.25) \quad A_n \triangle A = (A_n \setminus A) \cup (A \setminus A_n) \subseteq (B_n \setminus A) \cup (A \setminus C_n) = B_n \setminus C_n,$$

and so

$$(79.26) \quad d(A_n, A) \leq \mu(B_n \setminus C_n) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Let us check that (79.22) holds when $\sum_{j=1}^{\infty} d(A_{j+1}, A_j)$ converges. The main point is that

$$(79.27) \quad B_n \setminus A_n \subseteq \bigcup_{j=n}^{\infty} (A_{j+1} \setminus A_j), \quad A_n \setminus C_n \subseteq \bigcup_{j=n}^{\infty} (A_j \setminus A_{j+1})$$

for each n . More precisely, if $x \in B_n \setminus A_n$, then $x \in A_{j+1}$ for some $j \geq n+1$, and $x \notin A_n$. If j is the smallest integer such that $j \geq n$ and $x \in A_{j+1}$, then $x \notin A_j$, and so $x \in A_{j+1} \setminus A_j$, as desired. Similarly, if $y \in A_n \setminus C_n$, then $y \notin A_{j+1}$ for some $j \geq n$. If j is the smallest integer such that $j \geq n$ and $y \notin A_{j+1}$, then $y \in A_j$, and so $y \in A_j \setminus A_{j+1}$. This proves (79.27).

It follows that

$$(79.28) \quad \mu(B_n \setminus A_n) \leq \sum_{j=n}^{\infty} \mu(A_{j+1} \setminus A_j), \quad \mu(A_n \setminus C_n) \leq \sum_{j=n}^{\infty} \mu(A_j \setminus A_{j+1})$$

for each n . Hence

$$(79.29) \quad \mu(B_n \setminus C_n) = \mu(B_n \setminus A_n) + \mu(A_n \setminus C_n) \leq \sum_{j=n}^{\infty} d(A_{j+1}, A_j),$$

using the fact that $C_n \subseteq A_n \subseteq B_n$ in the first step. If $\sum_{j=1}^{\infty} d(A_{j+1}, A_j)$ converges, then the right side tends to 0 as $n \rightarrow \infty$, and so (79.22) holds. This implies that there is an $A \in \mathcal{A}$ such that $\lim_{n \rightarrow \infty} d(A_n, A) = 0$, by the earlier remarks. If instead $\{A_j\}_{j=1}^{\infty}$ satisfies the Cauchy condition

$$(79.30) \quad \lim_{j, l \rightarrow \infty} d(A_j, A_l) = 0,$$

then there is a subsequence $\{A_{j_n}\}_{n=1}^{\infty}$ of $\{A_j\}_{j=1}^{\infty}$ such that $\sum_{n=1}^{\infty} d(A_{j_{n+1}}, A_{j_n})$ converges. This implies that there is an $A \in \mathcal{A}$ such that $\lim_{n \rightarrow \infty} d(A_{j_n}, A) = 0$, as before. Using the Cauchy condition, one can check that $\lim_{j \rightarrow \infty} d(A_j, A) = 0$.

If $\mathcal{E} \subseteq \mathcal{A}$, then let $\overline{\mathcal{E}}$ be the collection of $A \in \mathcal{A}$ such that for each $\epsilon > 0$ there is an $E \in \mathcal{E}$ that satisfies $d(A, E) < \epsilon$. This is basically the same as the closure of a set in a metric space, except that $d(A, B)$ is only a semimetric. In particular, note that $\overline{\mathcal{E}}$ automatically contains every $A \in \mathcal{A}$ for which there is an $E \in \mathcal{E}$ such that $d(A, E) = 0$. As in the context of metric spaces, one can check that

$$(79.31) \quad \overline{\overline{\mathcal{E}}} = \overline{\mathcal{E}}.$$

If \mathcal{E} is a subalgebra of \mathcal{A} , then it is easy to see that $\overline{\mathcal{E}}$ is also a subalgebra of \mathcal{A} , using the properties of the distance related to unions, intersections, and complements discussed earlier in this section.

Let us check that $\overline{\mathcal{E}}$ is actually a σ -algebra when \mathcal{E} is an algebra. It suffices to show that $\bigcup_{j=1}^{\infty} A_j \in \overline{\mathcal{E}}$ for every sequence A_1, A_2, \dots of elements of $\overline{\mathcal{E}}$. Of course, $\bigcup_{j=1}^n A_j \in \overline{\mathcal{E}}$ for each n , because $\overline{\mathcal{E}}$ is an algebra. We also know that $\bigcup_{j=1}^n A_j$ converges to $\bigcup_{j=1}^{\infty} A_j$ as $n \rightarrow \infty$ with respect to $d(A, B)$, because of monotonicity. It follows that $\bigcup_{j=1}^{\infty} A_j \in \overline{\mathcal{E}}$, by combining these two facts.

If $A \in \overline{\mathcal{E}}$, then there is a sequence $\{A_j\}_{j=1}^{\infty}$ of elements of \mathcal{E} such that $\sum_{j=1}^{\infty} d(A_j, A)$ converges. This implies that $\sum_{j=1}^{\infty} d(A_{j+1}, A_j)$ converges, by the triangle inequality. Thus $\{A_j\}_{j=1}^{\infty}$ converges to $\limsup_{j \rightarrow \infty} A_j$, $\liminf_{j \rightarrow \infty} A_j$ with respect to $d(A, B)$, by the earlier discussion, and A differs from these limits by sets of measure 0. In particular, $A \in \mathcal{E}$ when \mathcal{E} is a σ -subalgebra of \mathcal{A} that contains all elements of \mathcal{A} with measure 0. It follows that $\overline{\mathcal{E}} = \mathcal{E}$ when \mathcal{E} is a σ -subalgebra of \mathcal{A} that contains the sets of measure 0.

80 Sequences of σ -subalgebras

Let (X, \mathcal{A}, μ) be a probability space, and let $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \dots$ be an increasing sequence of σ -subalgebras of \mathcal{A} . Thus $\mathcal{E} = \bigcup_{j=1}^{\infty} \mathcal{B}_j$ is a subalgebra of \mathcal{A} , but not necessarily a σ -subalgebra. If $\mathcal{C} = \overline{\mathcal{E}}$ is the closure of \mathcal{E} with respect to the semimetric $d(A, B)$, then \mathcal{C} is the smallest σ -subalgebra of \mathcal{A} that contains \mathcal{E} and the sets of measure 0, as in the previous section.

Put

$$(80.1) \quad f_n = f_{\mathcal{B}_n} = E(f \mid \mathcal{B}_n)$$

for each $f \in L^1(X, \mathcal{A})$ and $n \geq 1$, and

$$(80.2) \quad f_{\infty} = f_{\mathcal{C}} = E(f \mid \mathcal{C}).$$

Note that

$$(80.3) \quad f_n = E(f_\infty \mid \mathcal{B}_n)$$

for each n , since $\mathcal{B}_n \subseteq \mathcal{C}$. If $f \in L^p(X, \mathcal{A})$ for some p , $1 \leq p \leq \infty$, then $f_n \in L^p(X, \mathcal{B}_n)$ for each n , $f_\infty \in L^p(X, \mathcal{C})$, and

$$(80.4) \quad \|f_n\|_p \leq \|f_\infty\|_p \leq \|f\|_p.$$

If f happens to be measurable with respect to \mathcal{B}_l for some $l \geq 1$, then

$$(80.5) \quad f_n = f_\infty = f$$

for every $n \geq l$.

If $1 \leq p < \infty$, then

$$(80.6) \quad \bigcup_{l=1}^{\infty} L^p(X, \mathcal{B}_l)$$

is dense in $L^p(X, \mathcal{C})$. To see this, one can first approximate elements of $L^p(X, \mathcal{C})$ by simple functions that are measurable with respect to \mathcal{C} . The latter can then be approximated by simple functions that are measurable with respect to \mathcal{B}_l for some l , using the definition of \mathcal{C} . This implies that

$$(80.7) \quad \lim_{n \rightarrow \infty} f_n = f_\infty$$

in the L^p norm when $f \in L^p(X, \mathcal{A})$, $1 \leq p < \infty$. More precisely, one may as well take $f = f_\infty$, so that f is already measurable with respect to \mathcal{C} . If f is measurable with respect to \mathcal{B}_l for some l , then one can apply (80.5). Otherwise, one can approximate f by $g \in L^p(X, \mathcal{B}_l)$ for some l , by previous remarks about density in $L^p(X, \mathcal{C})$. The main point is that f_n is also approximated by g when $n \geq l$, uniformly in n , because of (80.4).

Suppose that X_1, X_2, \dots is a sequence of compact Hausdorff spaces, and that $X = \prod_{j=1}^{\infty} X_j$ is their Cartesian product, with the product topology. Let μ_j be a regular Borel probability measure on X_j for each j , and let μ be the corresponding product measure on X . Also let \mathcal{B}_n be the collection of subsets of X of the form $B \times \prod_{j=n+1}^{\infty} X_j$, where B is a Borel set in $\prod_{j=1}^n X_j$. If f is a continuous real or complex-valued function on X , then f_n is the function of x_1, \dots, x_n obtained by integrating f in the variables x_j for $j \geq n+1$. In this case, $\{f_n\}_{n=1}^{\infty}$ converges to f uniformly on X , because of the uniform continuity properties discussed in Section 72.

81 Martingales

Let (X, \mathcal{A}, μ) be a probability space, and let $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \dots$ be an increasing sequence of σ -subalgebras of \mathcal{A} , also known as a *filtration*. A sequence $\{f_j\}_{j=1}^{\infty}$ of functions on X is said to be a *martingale* with respect to this filtration if $f_j \in L^1(X, \mathcal{B}_j)$ for each j , and

$$(81.1) \quad f_j = E(f_l \mid \mathcal{B}_j)$$

when $1 \leq j \leq l$. In particular, this implies that

$$(81.2) \quad \|f_j\|_1 \leq \|f_l\|_1$$

for each $j \leq l$. If $f_j \in L^p(X, \mathcal{B}_j)$ for some p , $1 \leq p \leq \infty$, and every j , then

$$(81.3) \quad \|f_j\|_p \leq \|f_l\|_p$$

for each $j \leq l$. If $f \in L^1(X, \mathcal{A})$ and $f_j = E(f \mid \mathcal{B}_j)$ for each j , then $\{f_j\}_{j=1}^\infty$ is a martingale.

Let $(X_1, \mathcal{A}_1, \mu_1), (X_2, \mathcal{A}_2, \mu_2), \dots$ be a sequence of probability spaces, and let $X = \prod_{j=1}^\infty X_j$ be their Cartesian product, with the product measure μ on the corresponding σ -algebra \mathcal{A} . Also let \mathcal{B}_n be the collection of subsets of X of the form $B \times \prod_{j=n+1}^\infty X_j$, where B is a measurable subset of $\prod_{j=1}^n X_j$. This defines an increasing sequence of σ -subalgebras of \mathcal{A} . Let a_j be an integrable function on X_j such that

$$(81.4) \quad \int_{X_j} a_j d\mu_j = 0$$

for each j , which can also be considered as an integrable function on X that does not depend on x_l when $j \neq l$. In this case,

$$(81.5) \quad f_n = \sum_{j=1}^n a_j$$

defines a martingale with respect to this filtration.

Let (X, \mathcal{A}, μ) be any probability space again, with an increasing sequence \mathcal{B}_j of σ -algebras of \mathcal{A} . Also let $\{f_j\}_{j=1}^\infty$ be a martingale with respect to this filtration, with $f_j \in L^2(X, \mathcal{B}_j)$ for each j . Thus

$$(81.6) \quad \int_B f_j d\mu = \int_B f_{j+1} d\mu$$

for each $B \in \mathcal{B}_j$, which implies that

$$(81.7) \quad \int_X b f_j d\mu = \int_X b f_{j+1} d\mu$$

for every $b \in L^2(X, \mathcal{B}_j)$. Equivalently,

$$(81.8) \quad \int_X b (f_j - f_{j+1}) d\mu = 0$$

for every $b \in L^2(X, \mathcal{B}_j)$. It follows that the functions f_1 and $f_{j+1} - f_j$, $j \geq 1$, are all orthogonal to each other in $L^2(X, \mathcal{A})$.

82 L^p Boundedness

Let (X, \mathcal{A}, μ) be a probability space, and let $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \dots$ be an increasing sequence of σ -subalgebras of \mathcal{A} . As before, put $\mathcal{E} = \bigcup_{j=1}^{\infty} \mathcal{B}_j$, and let $\mathcal{C} = \overline{\mathcal{E}}$ be the closure of \mathcal{E} with respect to the semimetric $d(A, B)$. Let $1 < p \leq \infty$ be given, and let $\{f_j\}_{j=1}^{\infty}$ be a martingale on X with respect to the \mathcal{B}_j 's such that $f_j \in L^p(X, \mathcal{B}_j)$ for each j , and the L^p norms $\|f_j\|_p$ are uniformly bounded.

If $B \in \mathcal{B}_l$ for some l , then

$$(82.1) \quad \int_B f_l d\mu = \int_B f_n d\mu$$

when $n \geq l$. This implies that

$$(82.2) \quad \int_X f_l g d\mu = \int_X f_n g d\mu$$

when $g \in L^q(X, \mathcal{B}_l)$, where $1/p + 1/q = 1$. In particular,

$$(82.3) \quad \lim_{n \rightarrow \infty} \int_X f_n g d\mu$$

exists for every $g \in L^q(X, \mathcal{B}_l)$, $l \geq 1$. Note that $\bigcup_{l=1}^{\infty} L^q(X, \mathcal{B}_l)$ is dense in $L^q(X, \mathcal{C})$, as in Section 80, because $1 \leq q < \infty$. It follows that the limit (82.3) exists for every $g \in L^q(X, \mathcal{C})$, using also the uniform boundedness of the L^p norms of the f_j 's, as in Section 52.

More precisely, (82.3) defines a bounded linear functional on $L^q(X, \mathcal{C})$ under these conditions. The Riesz representation theorem implies that there is an $f \in L^p(X, \mathcal{C})$ such that

$$(82.4) \quad \lim_{n \rightarrow \infty} \int_X f_n g d\mu = \int_X f g d\mu$$

for every $g \in L^q(X, \mathcal{C})$ under these conditions. If $g \in L^q(X, \mathcal{B}_l)$ for some l , then we get that

$$(82.5) \quad \int_X f_l g d\mu = \int_X f g d\mu.$$

In particular,

$$(82.6) \quad \int_B f_l d\mu = \int_B f d\mu$$

for each $B \in \mathcal{B}_l$, which implies that

$$(82.7) \quad f_l = E(f \mid \mathcal{B}_l)$$

for each l .

If $1 < p < \infty$, then it follows that $\{f_l\}_{l=1}^{\infty}$ converges to f in the L^p norm, as in Section 80. If $p = 2$, then

$$(82.8) \quad \|f_n\|_2^2 = \|f_1\|_2^2 + \sum_{j=1}^{n-1} \|f_{j+1} - f_j\|_2^2$$

for each n , because of orthogonality, as in the previous section. The boundedness of the L^2 norms $\|f_n\|_2$ is equivalent to the convergence of the series

$$(82.9) \quad \sum_{j=1}^{\infty} \|f_{j+1} - f_j\|_2^2,$$

which implies the convergence of the series $\sum_{j=1}^{\infty} (f_{j+1} - f_j)$ in $L^2(X, \mathcal{C})$. This gives a more direct proof of the convergence of $\{f_j\}_{j=1}^{\infty}$ in $L^2(X, \mathcal{C})$ in this case. Of course, if $\{f_j\}_{j=1}^{\infty}$ is a martingale such that $f_j \in L^p(X, \mathcal{B}_l)$ converges to $f \in L^p(X, \mathcal{A})$ in the L^p norm for any p , $1 \leq p \leq \infty$, then $f \in L^p(X, \mathcal{C})$ and $f_l = E(f | \mathcal{B}_l)$ for each l , for basically the same reasons as before.

83 Uniform integrability

Let (X, \mathcal{A}, μ) be a probability space, and let $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \dots$ be an increasing sequence of σ -subalgebras of \mathcal{A} . Also let $\{f_j\}_{j=1}^{\infty}$ be a martingale with respect to this filtration with bounded L^1 norms, so that there is a $C \geq 0$ such that

$$(83.1) \quad \|f_n\|_1 \leq C$$

for each n . Note that this holds automatically when $f_j \geq 0$ for each j , because

$$(83.2) \quad \|f_j\|_1 = \int_X f_j d\mu = \int_X f_1 d\mu$$

for each $j \geq 1$ in this case.

Suppose that the f_j 's are *uniformly integrable* as well, in the sense that for each $\epsilon > 0$ there is a $\delta > 0$ such that

$$(83.3) \quad \int_A |f_n| d\mu < \epsilon$$

for every $A \in \mathcal{A}$ with $\mu(A) < \delta$ and every $n \geq 1$. It is well known that this condition holds automatically for a single integrable function, by approximating that function by bounded functions in the L^1 norm, for instance. Similarly, any finite collection of integrable functions has this property. Using this, it is easy to check that a sequence of integrable functions that converges in the L^1 norm is uniformly integrable. If there is a $p > 1$ such that $f_n \in L^p$ for each n and $\|f_n\|_p$ is uniformly bounded, then $\{f_n\}_{n=1}^{\infty}$ is uniformly integrable, because of Hölder's inequality.

If $\{f_n\}_{n=1}^{\infty}$ satisfies (83.1), then

$$(83.4) \quad \mu(\{x \in X : |f_n(x)| > t\}) \leq t^{-1} C$$

for each $t > 0$, by Tchebychev's inequality. If $\{f_j\}_{j=1}^{\infty}$ is uniformly integrable too, then it follows that

$$(83.5) \quad \int_{\{x \in X : |f_n(x)| > t\}} |f_n(x)| d\mu(x) \rightarrow 0 \quad \text{as } t \rightarrow \infty,$$

uniformly in n . Conversely, the latter condition implies that $\{f_n\}_{n=1}^\infty$ has bounded L^1 norms and is uniformly integrable.

As usual, put $\mathcal{E} = \bigcup_{j=1}^\infty \mathcal{B}_j$, and let $\mathcal{C} = \overline{\mathcal{E}}$ be the closure of \mathcal{E} with respect to the semimetric $d(A, B)$. Note that

$$(83.6) \quad \int_A f_n d\mu = \int_A f_l d\mu$$

for every $A \in \mathcal{B}_l$ and $n \geq l$. We would like to show that

$$(83.7) \quad \left\{ \int_A f_n d\mu \right\}_{n=1}^\infty$$

is a Cauchy sequence in \mathbf{R} or \mathbf{C} , as appropriate, for every $A \in \mathcal{C}$, and hence converges. This is obvious when $A \in \mathcal{E}$, and one can deal with $A \in \mathcal{C}$ by approximation, using uniform integrability. The main point is that

$$(83.8) \quad \int_A f_n d\mu, n \in \mathbf{Z}_+,$$

is an equicontinuous family of functions of $A \in \mathcal{A}$ with respect to the semimetric $d(A, B)$, since

$$(83.9) \quad \left| \int_A f_n d\mu - \int_B f_n d\mu \right| \leq \int_{A \Delta B} |f_n| d\mu$$

for every $A, B \in \mathcal{A}$.

Put

$$(83.10) \quad \nu(A) = \lim_{n \rightarrow \infty} \int_A f_n d\mu$$

for each $A \in \mathcal{C}$. Uniform integrability implies that for each $\epsilon > 0$ there is a $\delta > 0$ such that

$$(83.11) \quad |\nu(A)| \leq \epsilon$$

for every $A \in \mathcal{C}$ such that $\mu(A) < \delta$. This follows by taking the limit as $n \rightarrow \infty$ in the definition of uniform integrability of $\{f_n\}_{n=1}^\infty$, using the same δ as before.

Clearly $\nu(A)$ is finitely additive on \mathcal{C} , and countable additivity follows from this continuity condition. For if A_1, A_2, \dots is a sequence of pairwise-disjoint subsets of X in \mathcal{C} , then countable additivity of μ implies that

$$(83.12) \quad \lim_{k \rightarrow \infty} \mu\left(\bigcup_{j=k+1}^\infty A_j\right) = 0,$$

and hence

$$(83.13) \quad \lim_{k \rightarrow \infty} \nu\left(\bigcup_{j=k+1}^\infty A_j\right) = 0$$

too, by the continuity condition. Because of finite additivity, we also have that

$$(83.14) \quad \nu\left(\bigcup_{j=1}^\infty A_j\right) = \sum_{j=1}^k \nu(A_j) + \nu\left(\bigcup_{j=k+1}^\infty A_j\right)$$

for each $k \geq 1$. It follows that $\sum_{j=1}^{\infty} \nu(A_j)$ converges to $\nu\left(\bigcup_{j=1}^{\infty} A_j\right)$, as desired.

Thus ν is a countably-additive real or complex measure on \mathcal{C} , as appropriate. Moreover, ν is absolutely continuous with respect to the restriction of μ to \mathcal{C} . The Radon–Nikodym theorem implies that there is an $f \in L^1(X, \mathcal{C})$ such that

$$(83.15) \quad \nu(A) = \int_A f \, d\mu$$

for every $A \in \mathcal{C}$. In particular,

$$(83.16) \quad \int_A f_l \, d\mu = \int_A f \, d\mu$$

when $A \in \mathcal{B}_l$, which implies that

$$(83.17) \quad f_l = E(f \mid \mathcal{B}_l)$$

for each l . Conversely, this implies that $\{f_l\}_{l=1}^{\infty}$ converges to f in the L^1 norm, as in Section 80, which implies that $\{f_l\}_{l=1}^{\infty}$ is uniformly integrable.

84 Maximal functions, 3

Let (X, \mathcal{A}, μ) be a probability space, let $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \cdots$ be an increasing sequence of σ -subalgebras of \mathcal{A} , and let $\{f_j\}_{j=1}^{\infty}$ be a martingale on X with respect to this filtration. Consider the maximal functions

$$(84.1) \quad f_n^*(x) = \max_{1 \leq j \leq n} |f_j(x)|$$

and

$$(84.2) \quad f^*(x) = \sup_{j \geq 1} |f_j(x)|.$$

Note that f_n^* is measurable with respect to \mathcal{B}_n , and that

$$(84.3) \quad f^*(x) = \lim_{n \rightarrow \infty} f_n^*(x)$$

is measurable with respect to the smallest σ -algebra \mathcal{B}_{∞} that contains $\mathcal{E} = \bigcup_{j=1}^{\infty} \mathcal{B}_j$. If $X = [0, 1)$, μ is Lebesgue measure, and \mathcal{B}_j consists of unions of dyadic intervals of length 2^{-j} , then this is a variant of the dyadic maximal function, as in Section 58.

Put

$$(84.4) \quad E(t) = \{x \in X : f^*(x) > t\}$$

for each $t > 0$, as well as

$$(84.5) \quad E_1(t) = \{x \in X : |f_1(x)| > t\}$$

and

$$(84.6) \quad E_l(t) = \{x \in X : |f_l(x)| > t, f_{l-1}^*(x) \leq t\}$$

when $l \geq 2$. Thus $E_l(t) \in \mathcal{B}_l$ for each l, t , $E_l(t) \cap E_n(t) = \emptyset$ when $l < n$, and

$$(84.7) \quad E(t) = \bigcup_{l=1}^{\infty} E_l(t).$$

Similarly,

$$(84.8) \quad \bigcup_{l=1}^n E_l(t) = \{x \in X : f_n^*(x) > t\}$$

for each $n \geq 1$. If $l \leq n$, then

$$(84.9) \quad t \mu(E_l(t)) \leq \int_{E_l(t)} |f_l| d\mu \leq \int_{E_l(t)} |f_n| d\mu,$$

because $f_l = E(f_n | \mathcal{B}_l)$ and hence $|f_l| \leq E(|f_n| | \mathcal{B}_l)$, as in Section 78. This implies that

$$(84.10) \quad \begin{aligned} t \mu\left(\bigcup_{l=1}^n E_l(t)\right) &= \sum_{l=1}^n t \mu(E_l(t)) \leq \sum_{l=1}^n \int_{E_l(t)} |f_n| d\mu \\ &= \int_{\bigcup_{l=1}^n E_l(t)} |f_n| d\mu. \end{aligned}$$

Suppose now that the f_n 's have bounded L^1 norms, so that

$$(84.11) \quad \|f_n\|_1 \leq C$$

for some $C \geq 0$ and every $n \geq 1$. The previous estimate implies that

$$(84.12) \quad t \mu\left(\bigcup_{l=1}^n E_l(t)\right) \leq C$$

for each n . Hence

$$(84.13) \quad t \mu(E(t)) = t \mu\left(\bigcup_{l=1}^{\infty} E_l(t)\right) \leq C.$$

This is basically the same as the estimates in Sections 46 and 58, except that the measure μ here corresponds to Lebesgue measure before, and the martingale $\{f_j\}_{j=1}^{\infty}$ corresponds to the measure μ or function f before. The martingale may be generated by a function or measure on X , through conditional expectation.

We also have that

$$(84.14) \quad \begin{aligned} \sum_{l=1}^n \int_{E_l(t)} |f_l| d\mu &\leq \sum_{l=1}^n \int_{E_l(t)} |f_n| d\mu \\ &= \int_{\bigcup_{l=1}^n E_l(t)} |f_n| d\mu \leq C \end{aligned}$$

for each n , since $|f_l| \leq E(|f_n| \mid \mathcal{B}_l)$ when $l \leq n$. Hence

$$(84.15) \quad \sum_{l=1}^{\infty} \int_{E_l(t)} |f_l| d\mu \leq C.$$

This shows that the function h defined on X by $h = f_l$ on $E_l(t)$, $h = 0$ on $X \setminus E(t)$, is integrable, with $\|h\|_1 \leq C$.

85 Convergence almost everywhere

Let (X, \mathcal{A}, μ) be a probability space, let $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \cdots$ be an increasing sequence of σ -subalgebras of \mathcal{A} , and let $\{f_j\}_{j=1}^{\infty}$ be a martingale on X with respect to this filtration. Observe that

$$(85.1) \quad \{f_n - f_l\}_{n=l}^{\infty}$$

is a martingale with respect to the filtration $\mathcal{B}_l \subseteq \mathcal{B}_{l+1} \subseteq \cdots$ for each $l \geq 1$. Put

$$(85.2) \quad A_l(t) = \left\{ x \in X : \sup_{n \geq l} |f_n(x) - f_l(x)| > t \right\}$$

for every $l \geq 1$ and $t > 0$.

If $\|f_n\|_1$ is bounded, then

$$(85.3) \quad t \mu(A_l(t)) \leq \sup_{n \geq l} \|f_n - f_l\|_1$$

for every $t > 0$, as in the previous section. This implies that

$$(85.4) \quad t \mu\left(\bigcap_{l=1}^{\infty} A_l(t)\right) \leq \inf_{l \geq 1} \left(\sup_{n \geq l} \|f_n - f_l\|_1 \right),$$

for each $t > 0$, and hence

$$(85.5) \quad \mu\left(\bigcap_{l=1}^{\infty} A_l(t)\right) = 0$$

for every $t > 0$ when $\{f_n\}_{n=1}^{\infty}$ is a Cauchy sequence in $L^1(X, \mathcal{A})$. Thus

$$(85.6) \quad \mu\left(\bigcup_{k=1}^{\infty} \bigcap_{l=1}^{\infty} A_l(1/k)\right) = 0.$$

Of course,

$$(85.7) \quad X \setminus \left(\bigcup_{k=1}^{\infty} \bigcap_{l=1}^{\infty} A_l(1/k) \right) = \bigcap_{k=1}^{\infty} \bigcup_{l=1}^{\infty} (X \setminus A_l(1/k)).$$

If x is in this set, then it is easy to see that $\{f_n(x)\}_{n=1}^{\infty}$ is a Cauchy sequence in \mathbf{R} or \mathbf{C} , as appropriate. It follows that $\{f_n\}_{n=1}^{\infty}$ converges pointwise almost everywhere on X when it converges in the L^1 norm. As in Section 80, this

happens when there is an $f \in L^1(X, \mathcal{A})$ such that $f_n = E(f \mid \mathcal{B}_n)$ for each n . In particular, this happens when $\{f_n\}_{n=1}^\infty$ is uniformly integrable, as in Section 83. This includes the case where there is a $p > 1$ such that $f_n \in L^p(X, \mathcal{B}_n)$ for each n and $\|f_n\|_p$ is bounded, as in Section 82.

Suppose that we simply know that $\|f_n\|_1$ is uniformly bounded in n . Let $t > 0$ be given, and put $g_1 = f_1$, and

$$(85.8) \quad \begin{aligned} g_n(x) &= f_n(x) \quad \text{when } x \in X \setminus \left(\bigcup_{l=1}^{n-1} E_l(t) \right) \\ &= f_l(x) \quad \text{when } x \in E_l(t), 1 \leq l \leq n-1 \end{aligned}$$

for $n \geq 2$, where $E_l(t)$ is as in the previous section. Note that g_n is measurable with respect to \mathcal{B}_n for each n , because $E_l(t) \in \mathcal{B}_l \subseteq \mathcal{B}_n$ when $l \leq n$, as in the previous section, and f_l is measurable with respect to \mathcal{B}_l and hence \mathcal{B}_n when $l \leq n$. Moreover,

$$(85.9) \quad \begin{aligned} \int_X |g_n| d\mu &= \int_{X \setminus \left(\bigcup_{l=1}^n E_l(t) \right)} |f_n| d\mu + \sum_{l=1}^{n-1} \int_{E_l(t)} |f_l| d\mu \\ &\leq \int_{X \setminus \left(\bigcup_{l=1}^{n-1} E_l(t) \right)} |f_n| d\mu + \sum_{l=1}^{n-1} \int_{E_l(t)} |f_n| d\mu \end{aligned}$$

when $n \geq 2$, using the fact that $|f_l| \leq E(|f_n| \mid \mathcal{B}_l)$ in the second step. This implies that

$$(85.10) \quad \int_X |g_n| d\mu \leq \int_X |f_n| d\mu,$$

which obviously holds when $n = 1$ as well.

Let us check that $\{g_n\}_{n=1}^\infty$ is a martingale on X with respect to the \mathcal{B}_n 's. It suffices to show that

$$(85.11) \quad \int_A g_n d\mu = \int_A g_{n+1} d\mu$$

for each $A \in \mathcal{B}_n$ and $n \geq 1$, so that $g_n = E(g_{n+1} \mid \mathcal{B}_n)$. If $A \subseteq X \setminus \left(\bigcup_{l=1}^n E_l(t) \right)$, then $g_n = f_n$ and $g_{n+1} = f_{n+1}$ on A , and so

$$(85.12) \quad \int_A g_n d\mu = \int_A f_n d\mu = \int_A f_{n+1} d\mu = \int_A g_{n+1} d\mu.$$

This uses the facts that $f_n = E(f_{n+1} \mid \mathcal{B}_n)$ and $A \in \mathcal{B}_n$ in the middle step. If $A \subseteq E_n(t)$, then $g_n = f_n$ on A because $A \subseteq X \setminus \left(\bigcup_{l=1}^{n-1} E_l(t) \right)$, and $g_{n+1} = f_n$ on A by definition of g_{n+1} . Hence

$$(85.13) \quad \int_A g_n d\mu = \int_A f_n d\mu = \int_A g_{n+1} d\mu.$$

Similarly, if $A \subseteq E_l(t)$ for some $l = 1, \dots, n-1$, then $g_n = g_{n+1} = f_l$ on A , and so

$$(85.14) \quad \int_A g_n d\mu = \int_A f_l d\mu = \int_A g_{n+1} d\mu.$$

Every $A \in \mathcal{B}_n$ can be expressed as the disjoint union of its intersections with $X \setminus \left(\bigcup_{l=1}^n E_l(t) \right)$ and $E_l(t)$, $1 \leq l \leq n$, each of which is in \mathcal{B}_n . Thus (85.11) follows by combining the previous cases.

Now let us check that $\{g_n\}_{n=1}^\infty$ is uniformly integrable. Let h be the function on X defined by $h = f_n$ on $E_n(t)$ and $h = 0$ on $X \setminus E(t)$, as in the previous section. Observe that $g_n = h$ on $\bigcup_{l=1}^n E_l(t)$, while $g_n = f_n$ on $X \setminus \left(\bigcup_{l=1}^n E_l(t) \right)$. Moreover,

$$(85.15) \quad |g_n| = |f_n| \leq t$$

on $X \setminus \left(\bigcup_{l=1}^n E_l(t) \right)$, by definition of $E_l(t)$. This implies that

$$(85.16) \quad |g_n| \leq \max(|h|, t)$$

on X for each n , so that the uniform integrability of $\{g_n\}_{n=1}^\infty$ follows from the integrability of h .

Thus $\{g_n\}_{n=1}^\infty$ converges pointwise almost everywhere on X , as mentioned earlier in the section. By construction, $g_n = f_n$ on $X \setminus E(t)$ for each n , and so $\{f_n\}_{n=1}^\infty$ converges pointwise almost everywhere on $X \setminus E(t)$ for each $t > 0$. It follows that $\{f_n\}_{n=1}^\infty$ converges pointwise almost everywhere on

$$(85.17) \quad \bigcup_{k=1}^\infty (X \setminus E(k)) = X \setminus \left(\bigcap_{k=1}^\infty E(k) \right).$$

Of course,

$$(85.18) \quad \mu \left(\bigcap_{k=1}^\infty E(k) \right) \leq \inf_{k \geq 1} \mu(E(k)),$$

and $\mu(E(t)) \leq t^{-1} \sup_{n \geq 1} \|f_n\|_1 \rightarrow 0$ as $t \rightarrow \infty$, by (84.13). Hence

$$(85.19) \quad \mu \left(\bigcap_{k=1}^\infty E(k) \right) = 0,$$

which implies that $\{f_n\}_{n=1}^\infty$ converges pointwise almost everywhere on X .

86 Other measures

Let (X, \mathcal{A}, μ) be a probability space, and let $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \dots$ be an increasing sequence of σ -subalgebras of \mathcal{A} . Also let ν be a real or complex measure on a σ -algebra $\mathcal{B} \subseteq \mathcal{A}$ that contains each \mathcal{B}_j . Suppose that the restriction of ν to \mathcal{B}_j is absolutely continuous with respect to the restriction of μ to \mathcal{B}_j for each j . In particular, this happens when each \mathcal{B}_j is associated to a partition of

X by finitely or countably many sets of positive μ -measure, as in Section 77. Under these conditions, the Radon–Nikodym theorem implies that there is an $f_j \in L^1(X, \mathcal{B}_j)$ for each $j \geq 1$ such that

$$(86.1) \quad \int_B f_j d\mu = \nu(B)$$

for every $B \in \mathcal{B}_j$.

By construction, $\{f_j\}_{j=1}^\infty$ is a martingale on X with respect to the \mathcal{B}_j 's. Moreover,

$$(86.2) \quad \int_X |f_j| d\mu \leq |\nu|(X)$$

for each j , where $|\nu|$ denotes the total variation measure associated to ν . As in Section 78, (86.2) basically corresponds to the statement that the total variation of the restriction of ν to \mathcal{B}_j is less than or equal to the restriction of $|\nu|$ to \mathcal{B}_j . If ν is absolutely continuous with respect to the restriction of μ to \mathcal{B} , so that there is an $f \in L^1(X, \mathcal{B})$ such that

$$(86.3) \quad \nu(B) = \int_B f d\mu$$

for every $B \in \mathcal{B}$, then $f_j = E(f \mid \mathcal{B}_j)$ for each j .

Put

$$(86.4) \quad d'(A, B) = \mu(A \triangle B) + |\nu|(A \triangle B)$$

for every $A, B \in \mathcal{B}$. This defines a semimetric on \mathcal{B} , as in Section 79, and the closure \mathcal{C}' of $\mathcal{E} = \bigcup_{j=1}^\infty \mathcal{B}_j$ with respect to $d'(A, B)$ is a σ -subalgebra of \mathcal{B} that contains \mathcal{E} . More precisely, \mathcal{C}' is the smallest σ -subalgebra of \mathcal{B} that contains \mathcal{E} and the sets $A \in \mathcal{B}$ such that $\mu(A) = |\nu|(A) = 0$. In particular, \mathcal{C}' contains the smallest σ -algebra \mathcal{B}_∞ that contains \mathcal{E} , and \mathcal{C}' is contained in the closure \mathcal{C} of \mathcal{E} with respect to $d(A, B) = \mu(A \triangle B)$.

Suppose that $\{f_j\}_{j=1}^\infty$ converges to a function $f \in L^1(X, \mathcal{B})$ in the L^1 norm. If $A \in \mathcal{B}_l$ for some l , so that

$$(86.5) \quad \int_A f_j d\mu = \int_A f_l d\mu = \nu(A)$$

when $j \geq l$, then

$$(86.6) \quad \int_A f d\mu = \lim_{j \rightarrow \infty} \int_A f_j d\mu = \nu(A).$$

Thus

$$(86.7) \quad \int_A f d\mu = \nu(A)$$

for every $A \in \mathcal{E}$, and hence for every $A \in \mathcal{C}'$, because both sides of the equation are continuous with respect to $d'(A, B)$. This uses the analogue of uniform integrability for the single integrable function f . It follows that the restriction of ν to $\mathcal{B}_\infty \subseteq \mathcal{C}'$ is absolutely continuous with respect to the restriction of μ to \mathcal{B}_∞ under these conditions.

87 Finitely-additive measures

Let (X, \mathcal{A}, μ) be a probability space, let $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \dots$ be an increasing sequence of σ -subalgebras of \mathcal{A} , and let $\{f_j\}_{j=1}^\infty$ be a martingale on X with respect to this filtration. If we put

$$(87.1) \quad \nu(A) = \int_A f_j d\mu$$

when $A \in \mathcal{B}_j$, then ν is well-defined on $\mathcal{E} = \bigcup_{j=1}^\infty \mathcal{B}_j$, because

$$(87.2) \quad \int_A f_j d\mu = \int_A f_l d\mu$$

when $A \in \mathcal{B}_l$ and $j \geq l$. It is easy to see that ν is finitely additive on \mathcal{E} .

Suppose that the f_j 's have bounded L^1 norms, so that there is a $C \geq 0$ with the property that $\|f_j\|_1 \leq C$ for every $j \geq 1$. Let A_1, \dots, A_n be finitely many pairwise-disjoint subsets of X that are contained in \mathcal{E} . Thus $A_1, \dots, A_n \in \mathcal{B}_l$ for some l , and hence

$$(87.3) \quad \begin{aligned} \sum_{k=1}^n |\nu(A_k)| &= \sum_{k=1}^n \left| \int_{A_k} f_l d\mu \right| \leq \sum_{k=1}^n \int_{A_k} |f_l| d\mu \\ &= \int_{\bigcup_{k=1}^n A_k} |f_l| d\mu \leq C. \end{aligned}$$

Conversely, if

$$(87.4) \quad \sum_{k=1}^n |\nu(A_k)| \leq C$$

for every collection of finitely many pairwise disjoint elements A_1, \dots, A_n of \mathcal{B}_j , then $\|f_j\|_1 \leq C$. If ν has an extension to a countably-additive real or complex measure on a σ -algebra that contains \mathcal{E} , then (87.4) holds for each j , with C equal to the total variation of the extension of ν on X .

For example, let X be $[0, 1)$ equipped with Lebesgue measure, and let \mathcal{B}_j be the collection of subsets of $[0, 1)$ that are unions of dyadic intervals of length 2^{-j} . In this case, \mathcal{E} is the algebra of subsets of $[0, 1)$ that can be expressed as the union of finitely many dyadic intervals. Put

$$(87.5) \quad \begin{aligned} f_j(x) &= 0 && \text{when } 0 \leq x < 1 - 2^{-j}, \\ &= 2^j && \text{when } 1 - 2^{-j} \leq x < 1. \end{aligned}$$

Thus

$$(87.6) \quad \int_I f_j(x) dx = 0$$

when $I = [l 2^{-j}, (l+1) 2^{-j})$, $0 \leq l \leq 2^j - 2$, and

$$(87.7) \quad \int_I f_j(x) dx = 1$$

when $I = [1 - 2^{-j}, 1)$, which corresponds to $l = 2^j - 1$. It is easy to see that $\{f_j\}_{j=1}^\infty$ is a martingale on $[0, 1)$ with respect to this filtration. The finitely-additive measure ν on \mathcal{E} is characterized by $\nu(I) = 1$ when I is a dyadic interval with 1 as an endpoint, and $\nu(I) = 0$ for every other dyadic interval I . Note that $\|f_j\|_1 = 1$ for each j , and that (87.4) holds with $C = 1$, as it should. If $I_j = [1 - 2^{-j}, 1)$, then $I_{j+1} \subseteq I_j$ and $\nu(I_j) = 1$ for each $j \geq 1$, but $\bigcap_{j=1}^\infty I_j = \emptyset$. Basically, this martingale corresponds to a Dirac mass at the point 1. Since 1 is not included as an element of $X = [0, 1)$, there is no countably-additive measure on X from which the martingale is obtained.

Let (X, \mathcal{A}, μ) be any probability space again, with an increasing sequence $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \dots$ of σ -subalgebras of \mathcal{A} , and let $\{f_j\}_{j=1}^\infty$ be a martingale on X with respect to this filtration with bounded L^1 norms. As in Section 85, $\{f_j\}_{j=1}^\infty$ converges pointwise almost everywhere on X . The limit determines an element g of $L^1(X, \mathcal{C})$, where \mathcal{C} is the closure of \mathcal{E} with respect to the usual semimetric $d(A, B) = \mu(A \triangle B)$ on \mathcal{A} , as in Section 79. Equivalently, \mathcal{C} is the smallest σ -subalgebra of \mathcal{A} that contains \mathcal{E} and every $A \in \mathcal{A}$ with $\mu(A) = 0$. If $g_j = E(g \mid \mathcal{B}_j)$ for each j , then $\{g_j\}_{j=1}^\infty$ is a martingale on X with respect to this filtration that converges to g in the L^1 norm, as in Section 80. Hence $\{g_j\}_{j=1}^\infty$ also converges to g pointwise almost everywhere on X , as in Section 85. If $h_j = f_j - g_j$, then $\{h_j\}_{j=1}^\infty$ is also a martingale on X with respect to this filtration, and with bounded L^1 norms. By construction, $\{h_j\}_{j=1}^\infty$ converges to 0 pointwise almost everywhere on X . One can think of $\{g_j\}_{j=1}^\infty$ as the “regular part” of the martingale $\{f_j\}_{j=1}^\infty$, and of $\{h_j\}_{j=1}^\infty$ as the “singular part” of $\{f_j\}_{j=1}^\infty$.

88 Maximal functions, 4

Let (X, \mathcal{A}, μ) be a probability space, and let $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \dots$ be an increasing sequence of σ -subalgebras of \mathcal{A} . If $f \in L^1(X, \mathcal{A})$, then $f_j = E(f \mid \mathcal{B}_j)$ defines a martingale on X with respect to this filtration, and we get the corresponding maximal function

$$(88.1) \quad f^*(x) = \sup_{j \geq 1} |f_j(x)|,$$

as before. Note that $f \mapsto f^*$ is sublinear, in the sense that

$$(88.2) \quad (af)^* = |a| f^*$$

and

$$(88.3) \quad (f + g)^* \leq f^* + g^*$$

for every $f, g \in L^1(X, \mathcal{A})$ and $a \in \mathbf{R}$ or \mathbf{C} .

If $f \in L^\infty(X, \mathcal{A})$, then $f_j \in L^\infty(X, \mathcal{A})$ for each j , and

$$(88.4) \quad \|f_j\|_\infty \leq \|f\|_\infty,$$

as in Section 78. This implies that $f^* \in L^\infty(X, \mathcal{A})$, and that

$$(88.5) \quad \|f^*\|_\infty \leq \|f\|_\infty.$$

If $f \in L^1(X, \mathcal{A})$, then

$$(88.6) \quad \|f_j\|_1 \leq \|f\|$$

for each j , as in Section 78. Put

$$(88.7) \quad E(t) = \{x \in X : f^*(x) > t\}$$

for each $t > 0$, so that

$$(88.8) \quad \mu(E(t)) \leq t^{-1} \|f\|_1,$$

as in Section 84.

Let g be the function defined on X by

$$(88.9) \quad \begin{aligned} g(x) &= f(x) && \text{when } |f(x)| \leq t/2 \\ &= 0 && \text{when } |f(x)| > t/2. \end{aligned}$$

Thus $g \in L^\infty(X, \mathcal{A})$, and hence $g^* \in L^\infty(X, \mathcal{A})$, with

$$(88.10) \quad \|g^*\|_\infty \leq \|g\|_\infty \leq \frac{t}{2}.$$

This implies that

$$(88.11) \quad f^*(x) \leq (f - g)^*(x) + g^*(x) \leq (f - g)^*(x) + \frac{t}{2}$$

for almost every $x \in X$, so that

$$(88.12) \quad (f - g)^*(x) > t/2$$

for almost every $x \in E(t)$.

It follows that

$$(88.13) \quad \mu(E(t)) \leq \mu(\{x \in X : (f - g)^*(x) > t/2\}) \leq t^{-1} \|f - g\|_1.$$

Using the definition of g , we get that

$$(88.14) \quad \mu(E(t)) \leq t^{-1} \int_{\{x \in X : |f(x)| > t/2\}} |f(x)| d\mu(x).$$

If h is a nonnegative measurable function on X , then

$$(88.15) \quad A(h) = \{(x, r) \in X \times \mathbf{R} : 0 < r < h(x)\}$$

is a measurable subset of $X \times \mathbf{R}$. This is easy to see when h is a measurable simple function, and otherwise h can be approximated by an increasing sequence of measurable simple functions. Integrating $p r^{p-1}$ over $A(h)$ with respect to the product of μ on X and Lebesgue measure on \mathbf{R} , we get that

$$(88.16) \quad \int_X h^p d\mu = \int_0^\infty p r^{p-1} \mu(\{x \in X : h(x) > r\}) dr.$$

More precisely, the left side of (88.16) obtained by integrating $p r^{p-1}$ over $A(h)$ in r and then x , while the right side is obtained by integrating in x and then r .

In particular, if $1 < p < \infty$, then

$$(88.17) \quad \begin{aligned} \int_X (f^*)^p d\mu &= \int_0^\infty p t^{p-1} \mu(E(t)) dt \\ &\leq \int_0^\infty p t^{p-2} \int_{\{x \in X : |f(x)| > t/2\}} |f(x)| d\mu(x) dt, \end{aligned}$$

by (88.14). Interchanging the order of integration, we get that

$$(88.18) \quad \begin{aligned} \int_X (f^*)^p d\mu &\leq \int_X \int_0^{2|f(x)|} |f(x)| p t^{p-2} dt d\mu(x) \\ &= \frac{p 2^{p-1}}{p-1} \int_X |f(x)|^p d\mu(x). \end{aligned}$$

This shows that $f^* \in L^p(X, \mathcal{A})$ when $f \in L^p(X, \mathcal{A})$ and $p > 1$.

By contrast, if $f \in L^p(X, \mathcal{A})$, then

$$(88.19) \quad (f^*(x))^p \leq (|f|^p)^*(x),$$

by (78.18). As before,

$$(88.20) \quad \mu(\{x \in X : (|f|^p)^*(x) > t\}) \leq t^{-1} \int_X |f(x)|^p d\mu(x)$$

for every $t > 0$. This implies that

$$(88.21) \quad \mu(\{x \in X : (f^*(x))^p > t\}) \leq t^{-1} \int_X |f(x)|^p d\mu(x),$$

or equivalently

$$(88.22) \quad \mu(\{x \in X : f^*(x) > t\}) \leq t^{-p} \int_X |f(x)|^p d\mu(x)$$

for every $t > 0$. This is not strong enough to imply that $f^* \in L^p$, by integrating over t as in the previous paragraph. However, it does have the advantage of working uniformly over $p \geq 1$.

Note that we get the same estimates for the dyadic maximal function, as in Section 58, which corresponds to $X = [0, 1)$ with Lebesgue measure, and where \mathcal{B}_j consists of unions of dyadic intervals of length 2^{-j} . There are also similar estimates for the Hardy–Littlewood maximal function on the real line, as in Section 46, but with an extra factor of 2 in (88.8), and in the later steps.

89 Decreasing sequences of σ -algebras

Let (X, \mathcal{A}, μ) be a probability space, and suppose that $\mathcal{A}_1 \supseteq \mathcal{A}_2 \supseteq \cdots$ is a decreasing sequence of σ -subalgebras of \mathcal{A} . As a basic scenario, it may be

that $X = \prod_{j=1}^{\infty} X_j$ is the Cartesian product of a sequence of probability spaces X_1, X_2, \dots , and that \mathcal{A}_n consists of subsets of X of the form $\prod_{j=1}^n X_j \times A$, where A is a measurable subset of $\prod_{j=n+1}^{\infty} X_j$. In this case, conditional expectation with respect to \mathcal{A}_n corresponds to integrating a function on X in x_1, \dots, x_n . Basically, conditional expectation with respect to smaller σ -algebras corresponds to averaging functions over larger sets.

Note that $\mathcal{A}_{\infty} = \bigcap_{j=1}^{\infty} \mathcal{A}_j$ is automatically a σ -subalgebra of \mathcal{A} . If $A_j \in \mathcal{A}_j$ satisfies $A_j \subseteq A_{j+1}$ for each j , then $\bigcup_{j=1}^{\infty} A_j \in \mathcal{A}_{\infty}$, because

$$(89.1) \quad \bigcup_{j=1}^{\infty} A_j = \bigcup_{j=n}^{\infty} A_j \in \mathcal{A}_n$$

for each n . Similarly, if $B_j \in \mathcal{A}_j$ satisfies $B_{j+1} \subseteq B_j$ for each j , then

$$(89.2) \quad \bigcap_{j=1}^{\infty} B_j = \bigcap_{j=n}^{\infty} B_j \in \mathcal{A}_n$$

for each n , and so $\bigcap_{j=1}^{\infty} B_j \in \mathcal{A}_{\infty}$. If $E_j \in \mathcal{A}_j$ for each j , then it follows that

$$(89.3) \quad \limsup_{j \rightarrow \infty} E_j = \bigcap_{l=1}^{\infty} \left(\bigcup_{j=l}^{\infty} E_j \right), \quad \liminf_{j \rightarrow \infty} E_j = \bigcup_{l=1}^{\infty} \left(\bigcap_{j=l}^{\infty} E_j \right)$$

are also elements of \mathcal{A}_{∞} , by taking $A_l = \bigcup_{j=l}^{\infty} E_j$ and $B_l = \bigcap_{j=l}^{\infty} E_j$.

If f is a measurable function on X with respect to \mathcal{A} , and if f_j is a measurable function on X with respect to \mathcal{A}_j such that $f = f_j$ almost everywhere for each j , then there is a measurable function f_{∞} on X with respect to \mathcal{A}_{∞} such that $f = f_{\infty}$ almost everywhere. To see this, put

$$(89.4) \quad E_j = \{x \in X : f_j(x) = f_{j+1}(x)\},$$

so that $E_j \in \mathcal{A}_j$ for each j . Thus $B_l = \bigcap_{j=l}^{\infty} E_j \in \mathcal{A}_l$, and $f_j(x) = f_l(x)$ for every $x \in B_l$ and $j \geq l$. By hypothesis, $\mu(X \setminus E_j) = 0$ for each j , and so $\mu(X \setminus B_l) = 0$ for each l , since $X \setminus B_l = \bigcup_{j=l}^{\infty} (X \setminus E_j)$. We also have that $\bigcup_{l=1}^{\infty} B_l \in \mathcal{A}_{\infty}$, as in the previous paragraph. Put

$$(89.5) \quad \begin{aligned} f_{\infty}(x) &= 0 && \text{when } x \in X \setminus \left(\bigcup_{l=1}^{\infty} B_l \right) \\ &= f_l(x) && \text{when } x \in B_l \text{ for some } l \geq 1. \end{aligned}$$

This is well defined, because $f_j(x) = f_l(x)$ when $x \in B_l$ and $j \geq l$. Moreover, f_{∞} is measurable with respect to \mathcal{A}_l for every l , because f_l is measurable with respect to \mathcal{A}_l . This implies that f_{∞} is measurable with respect to \mathcal{A}_{∞} . It is easy to see that $f = f_{\infty}$ almost everywhere, since $f = f_l$ almost everywhere.

Let $\{f_j\}_{j=1}^{\infty}$ be a sequence of real-valued functions on X such that f_j is measurable with respect to \mathcal{A}_j for each j . Thus

$$(89.6) \quad \sup_{j \geq l} f_j(x), \quad \inf_{j \geq l} f_j(x)$$

are measurable with respect to \mathcal{A}_l for each l . This implies that

$$(89.7) \quad \limsup_{j \rightarrow \infty} f_j(x), \quad \liminf_{j \rightarrow \infty} f_j(x)$$

are measurable with respect to \mathcal{A}_l for each l , and hence are measurable with respect to \mathcal{A}_∞ . In particular, the set of $x \in X$ on which $\{f_j(x)\}_{j=1}^\infty$ converges is measurable with respect to \mathcal{A}_∞ , and the limit defines a measurable function with respect to \mathcal{A}_∞ on this set. The analogous statement for complex-valued functions follows by considering the real and imaginary parts separately.

Let $f \in L^1(X, \mathcal{A})$ be given, and put $f_j = E(f \mid \mathcal{A}_j)$ for each $j \geq 1$, and $f_0 = f$. Thus

$$(89.8) \quad f = \sum_{j=1}^n (f_{j-1} - f_j) + f_n$$

for each $n \geq 1$. If $f \in L^2(X, \mathcal{A})$, then the functions $f_{j-1} - f_j$, $1 \leq j \leq n$, and f_n are pairwise orthogonal in $L^2(X, \mathcal{A})$, as in Section 81. This implies that

$$(89.9) \quad \|f\|_2^2 = \sum_{j=1}^n \|f_{j-1} - f_j\|_2^2 + \|f_n\|_2^2$$

for each n , and hence that $\sum_{j=1}^\infty \|f_{j-1} - f_j\|_2^2$ converges. Therefore

$$(89.10) \quad \sum_{j=1}^\infty (f_{j-1} - f_j)$$

converges in $L^2(X, \mathcal{A})$, by orthogonality, which implies that $\{f_n\}_{n=1}^\infty$ converges in $L^2(X, \mathcal{A})$. Of course, $\{f_n\}_{n=l}^\infty$ converges in $L^2(X, \mathcal{A}_l)$ for each l , and the limits correspond to the same element of $L^2(X, \mathcal{A})$ for each l . Thus the limit may be represented by an element f_∞ of $L^2(X, \mathcal{A}_\infty)$, by the earlier remarks. In particular, $f_\infty = E(f_\infty \mid \mathcal{A}_\infty)$, which implies that

$$(89.11) \quad f_\infty = E(f \mid \mathcal{A}_\infty).$$

This uses the fact that $E(f \mid \mathcal{A}_\infty) = E(f_j \mid \mathcal{A}_\infty)$ for each j , since $f_j = E(f \mid \mathcal{A}_j)$ and $\mathcal{A}_\infty \subseteq \mathcal{A}_j$, and the convergence of $\{f_j\}_{j=1}^\infty$ to f_∞ in $L^2(X, \mathcal{A})$.

If $f \in L^p(X, \mathcal{A})$, $1 \leq p < 2$, then $L^2(X, \mathcal{A})$ is a dense linear subspace of $L^p(X, \mathcal{A})$, and one can use this to show that $\{f_j\}_{j=1}^\infty$ converges to $E(f \mid \mathcal{A}_\infty)$ in the L^p norm. This also uses the fact that the conditional expectation operators have operator norm 1 on L^p for each p . If $f \in L^\infty(X, \mathcal{A})$, then $f_j \in L^\infty(X, \mathcal{A}_j)$ with $\|f_j\|_\infty \leq \|f\|_\infty$ for each j . This together with convergence in $L^2(X, \mathcal{A})$ implies convergence in $L^p(X, \mathcal{A})$ for every $p < \infty$. If $f \in L^p(X, \mathcal{A})$, $2 < p < \infty$, then one can show again that $\{f_j\}_{j=1}^\infty$ converges to $E(f \mid \mathcal{A}_\infty)$ in the L^p norm, since this holds on the dense linear subspace $L^\infty(X, \mathcal{A})$ of $L^p(X, \mathcal{A})$, and because the expectation operators are uniformly bounded on L^p .

There are also maximal function estimates in this context. To see this, one can begin by observing that

$$(89.12) \quad f_n^*(x) = \max_{1 \leq j \leq n} |f_j(x)|$$

is basically the same as before, because one can simply rearrange the indices to get an increasing sequence of n σ -algebras. Hence the estimates for f_n^* are the same as before, and the corresponding estimates for

$$(89.13) \quad f^*(x) = \sup_{j \geq 1} |f_j(x)|$$

can be obtained by passing to the limit as $n \rightarrow \infty$. Convergence almost everywhere then follows from convergence in the L^1 norm, as in Section 85.

90 Doubly-infinite sequences

A probability space (X, \mathcal{A}, μ) may also have a doubly-infinite sequence

$$(90.1) \quad \cdots \subseteq \mathcal{B}_{-1} \subseteq \mathcal{B}_0 \subseteq \mathcal{B}_1 \subseteq \cdots$$

of σ -subalgebras of \mathcal{A} . In particular, this occurs very naturally in the context of doubly-infinite products. Let $(X_j, \mathcal{A}_j, \mu_j)$, $j \in \mathbf{Z}$ be a family of probability spaces indexed by the integers, and let $X = \prod_{j=-\infty}^{\infty} X_j$ be their Cartesian product, equipped with the product measure μ . Thus X consists of the doubly-infinite sequences $x = \{x_j\}_{j=-\infty}^{\infty}$ such that $x_j \in X_j$ for each j . If \mathcal{B}_n is the collection of subsets of X of the form $A \times \prod_{j=n+1}^{\infty} X_j$, where A is a measurable subset of $\prod_{j=-\infty}^n X_j$, then \mathcal{B}_n is a σ -subalgebra of the σ -algebra of measurable subsets of X , and $\mathcal{B}_n \subseteq \mathcal{B}_{n+1}$ for each n .

Suppose that $(X_j, \mathcal{A}_j, \mu_j)$ is a copy of the same probability space for each j . In this case, we can define the shift mapping $T : X \rightarrow X$ by $T(x) = y$, where $x = \{x_j\}_{j=-\infty}^{\infty}$, $y = \{y_j\}_{j=-\infty}^{\infty} \in X$ satisfy

$$(90.2) \quad y_j = x_{j-1}$$

for each j . If $A \subseteq X$ is measurable, then $T(A)$ is also measurable, and

$$(90.3) \quad \mu(T(A)) = \mu(A).$$

Similarly, T maps \mathcal{B}_n onto \mathcal{B}_{n+1} for each n .

If the X_j 's are compact Hausdorff topological spaces, then X is too, with respect to the product topology. If the X_j 's are all copies of the same topological space, then T is a homeomorphism. If the X_j 's are all metrizable, then X is as well, as in Section 69. However, this does not mean that there is a metric $d(x, y)$ on X that determines the product topology and which is invariant under T in the sense that

$$(90.4) \quad d(T(x), T(y)) = d(x, y)$$

for every $x, y \in X$. If $x, y \in X$ satisfy $x_j = x_l$ for every $j, l \in \mathbf{Z}$ and $x_j = y_j$ for all but exactly one $j \in \mathbf{Z}$, then $T(x) = x$ and $\lim_{n \rightarrow \infty} T^n(y) = x$, which would not be possible if there were an invariant metric.

91 Submartingales

Let (X, \mathcal{A}, μ) be a probability space, and let $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \dots$ be an increasing sequence of σ -subalgebras of \mathcal{A} . Also let $\{f_j\}_{j=1}^\infty$ be a sequence of real-valued functions on X such that $f_j \in L^1(X, \mathcal{B}_j)$ for each j . We say that $\{f_j\}_{j=1}^\infty$ is a *submartingale* on X with respect to this filtration if

$$(91.1) \quad f_j \leq E(f_{j+1} \mid \mathcal{B}_j)$$

almost everywhere on X with respect to μ for each j . Similarly, $\{f_j\}_{j=1}^\infty$ is a *supermartingale* if

$$(91.2) \quad f_j \geq E(f_{j+1} \mid \mathcal{B}_j)$$

almost everywhere on X for each j . Thus $\{f_j\}_{j=1}^\infty$ is a martingale if and only if it is both a submartingale and a supermartingale, and $\{f_j\}_{j=1}^\infty$ is a supermartingale if and only if $\{-f_j\}_{j=1}^\infty$ is a submartingale.

If $\{g_j\}_{j=1}^\infty$ is a real or complex martingale on X with respect to the \mathcal{B}_j 's, then $\{|g_j|\}_{j=1}^\infty$ is a submartingale on X . If in addition $g_j \in L^p(X, \mathcal{B}_j)$ for some p , $1 < p < \infty$, and each j , then $\{|g_j|^p\}_{j=1}^\infty$ is a submartingale as well. More generally, if ϕ is a convex function on an interval I in the real line, which may be unbounded, and if g_j takes values in I and $\phi \circ g_j \in L^1(X, \mathcal{B}_j)$ for each j , then $\{\phi \circ g_j\}_{j=1}^\infty$ is a submartingale. These statements use the remarks in Section 78. The latter also works when $\{g_j\}_{j=1}^\infty$ is a submartingale and ϕ is both convex and monotone increasing on I .

If $\{f_j\}_{j=1}^\infty$ is a submartingale on X and a is a nonnegative real number, then $\{a + f_j\}_{j=1}^\infty$ is a submartingale. If $\{f_j\}_{j=1}^\infty, \{g_j\}_{j=1}^\infty$ are submartingales, then their sum $\{f_j + g_j\}_{j=1}^\infty$ is a martingale too. Their maximum $\{\max(f_j, g_j)\}_{j=1}^\infty$ is a submartingale as well, because

$$(91.3) \quad f_j \leq E(f_{j+1} \mid \mathcal{B}_j) \leq E(\max(f_{j+1}, g_{j+1}) \mid \mathcal{B}_j)$$

and

$$(91.4) \quad g_j \leq E(g_{j+1} \mid \mathcal{B}_j) \leq E(\max(f_{j+1}, g_{j+1}) \mid \mathcal{B}_j)$$

imply that

$$(91.5) \quad \max(f_j, g_j) \leq E(\max(f_{j+1}, g_{j+1}) \mid \mathcal{B}_j).$$

Of course, $\{f_j + g_j\}_{j=1}^\infty$ is a martingale when $\{f_j\}_{j=1}^\infty, \{g_j\}_{j=1}^\infty$ are martingales, but $\{\max(f_j, g_j)\}_{j=1}^\infty$ is not normally a martingale in this case.

Let $\{f_j\}_{j=1}^\infty$ be a sequence of real-valued functions on X with $f_j \in L^1(X, \mathcal{B}_j)$ for each j , as before. Thus $\{f_j\}_{j=1}^\infty$ is determined by the initial function f_1 and the sequence of differences $f_{j+1} - f_j$. The condition that $\{f_j\}_{j=1}^\infty$ be a martingale can be expressed by

$$(91.6) \quad E(f_{j+1} - f_j \mid \mathcal{B}_j) = 0$$

for each j , while the condition that $\{f_j\}_{j=1}^\infty$ be a submartingale is expressed by

$$(91.7) \quad E(f_{j+1} - f_j \mid \mathcal{B}_j) \geq 0.$$

Suppose that $\{f_j\}_{j=1}^\infty$ is a submartingale, and put

$$(91.8) \quad a_j = E(f_{j+1} - f_j \mid \mathcal{B}_j) \geq 0$$

for each j . Also put $A_l = \sum_{j=1}^{l-1} a_j$ when $l \geq 2$, and $A_1 = 0$. Note that $A_l \in L^1(X, \mathcal{B}_{l-1})$ when $l \geq 2$, and $A_l(x)$ is monotone increasing in l for each $x \in X$. By construction, $\{f_l - A_l\}_{l=1}^\infty$ is a martingale, because

$$(91.9) \quad (f_{l+1} - A_{l+1}) - (f_l - A_l) = f_{l+1} - f_l - a_l$$

and

$$(91.10) \quad \begin{aligned} E(f_{l+1} - f_l - a_l \mid \mathcal{B}_l) &= E(f_{l+1} - f_l \mid \mathcal{B}_l) - E(a_l \mid \mathcal{B}_l) \\ &= E(f_{l+1} - f_l \mid \mathcal{B}_l) - a_l = 0. \end{aligned}$$

Conversely, if $\{\phi_j\}_{j=1}^\infty$ is any sequence of real-valued functions on X such that $\phi_j \in L^1(X, \mathcal{B}_j)$ and $\phi_j \leq \phi_{j+1}$ for each j , then $\{\phi_j\}_{j=1}^\infty$ is a submartingale on X . If $\{\psi_j\}_{j=1}^\infty$ is a martingale on X , then $\{\phi_j + \psi_j\}_{j=1}^\infty$ is also a submartingale. Every submartingale on X can be represented in this way, by the remarks in the previous paragraph.

Suppose that $f_j = \phi_j + \psi_j$ is a submartingale on X , where $\{\psi_j\}_{j=1}^\infty$ is a martingale, and $\phi_j \leq \phi_{j+1}$ for each j . If the integrals

$$(91.11) \quad \int_X f_j d\mu$$

have an upper bound in \mathbf{R} , then the integrals

$$(91.12) \quad \int_X \phi_j d\mu$$

also have an upper bound in \mathbf{R} , because $\int_X \psi_j d\mu$ is constant in j , by hypothesis. This implies that $\{\phi_j\}_{j=1}^\infty$ converges pointwise almost everywhere on X and in the L^1 norm, by the monotone convergence theorem. In particular, the ϕ_j 's have bounded L^1 norms. If the f_j 's have bounded L^1 norms, then it follows that the ψ_j 's have bounded L^1 norms too. This implies that $\{\psi_j\}_{j=1}^\infty$ converges pointwise almost everywhere on X , as in Section 85, and hence that $\{f_j\}_{j=1}^\infty$ converges pointwise almost everywhere on X as well. Similarly, $\{\psi_j\}_{j=1}^\infty$ converges in the L^1 norm when $\{f_j\}_{j=1}^\infty$ converges in the L^1 norm. Conversely, $\{f_j\}_{j=1}^\infty$ converges in the L^1 norm when $\{\psi_j\}_{j=1}^\infty$ converges in the L^1 norm and the integrals (91.11) have an upper bound in \mathbf{R} . If $\{f_j\}_{j=1}^\infty$ is uniformly integrable, then $\{\psi_j\}_{j=1}^\infty$ is uniformly integrable, because $\{\phi_j\}_{j=1}^\infty$ converges in L^1 and hence is uniformly integrable. This implies that $\{\psi_j\}_{j=1}^\infty$ converges in L^1 too, as in Section 83, so that $\{f_j\}_{j=1}^\infty$ converges in L^1 as well, as in the case of martingales.

Let $\{f_j\}_{j=1}^\infty$ be a submartingale on X , and observe that

$$(91.13) \quad \int_X f_j d\mu \leq \int_X E(f_{j+1} \mid \mathcal{B}_j) d\mu = \int_X f_{j+1} d\mu$$

for each j . If $j \geq l$, then

$$(91.14) \quad E(f_j \mid \mathcal{B}_l) \leq E(E(f_{j+1} \mid \mathcal{B}_j) \mid \mathcal{B}_l) = E(f_{j+1} \mid \mathcal{B}_l),$$

and

$$(91.15) \quad \int_X (E(f_j \mid \mathcal{B}_l) - f_l) d\mu = \int_X f_j d\mu - \int_X f_l d\mu.$$

Suppose that $\int_X f_j d\mu$ has an upper bound in \mathbf{R} , and hence converges in \mathbf{R} , by monotonicity. The monotone convergence theorem implies that $\{E(f_j \mid \mathcal{B}_l)\}_{j=l}^\infty$ converges in $L^1(X, \mathcal{B}_l)$ for each l . It is easy to check that the limit g_l satisfies

$$(91.16) \quad g_l = E(g_{l+1} \mid \mathcal{B}_l)$$

for each l , because

$$(91.17) \quad E(E(f_j \mid \mathcal{B}_{l+1}) \mid \mathcal{B}_l) = E(f_j \mid \mathcal{B}_l)$$

for each j, l . Thus $\{g_l\}_{l=1}^\infty$ is a martingale, and

$$(91.18) \quad f_l \leq E(f_j \mid \mathcal{B}_l) \leq g_l$$

when $j \geq l$, by construction. Moreover,

$$(91.19) \quad \int_X g_l d\mu = \lim_{j \rightarrow \infty} \int_X E(f_j \mid \mathcal{B}_l) d\mu = \lim_{j \rightarrow \infty} \int_X f_j d\mu$$

for each l , which implies that

$$(91.20) \quad \lim_{l \rightarrow \infty} \int_X (g_l - f_l) d\mu = 0,$$

since $\int_X g_l d\mu$ is constant in l .

Conversely, if $\{g'_j\}_{j=1}^\infty$ is a martingale on X such that $f_j \leq g'_j$ for each j , then

$$(91.21) \quad \int_X f_j d\mu \leq \int_X g'_j d\mu$$

has an upper bound in \mathbf{R} , because $\int_X g'_j d\mu$ is constant in j . In addition,

$$(91.22) \quad E(f_j \mid \mathcal{B}_l) \leq E(g'_j \mid \mathcal{B}_l) = g'_l$$

when $j \geq l$, which implies that $g_l \leq g'_l$ for each l , where g_l is as in the preceding paragraph.

Let $\{f_j\}_{j=1}^\infty$ be a submartingale on X again, and put

$$(91.23) \quad f_n^*(x) = \max(f_1(x), \dots, f_n(x)).$$

This is a bit different from the situation for martingales discussed in Section 84, since we do not take the absolute values of the functions. However, if $\{g_j\}_{j=1}^\infty$ is a martingale, then $f_j = |g_j|$ is a submartingale, and

$$(91.24) \quad f_n^*(x) = \max(|g_1(x)|, \dots, |g_n(x)|)$$

is the same as before. Note that f_n^* is measurable with respect to \mathcal{B}_n , as before.

Put
(91.25)
$$A_n(t) = \{x \in X : f_n^*(x) > t\}$$

for each $n \geq 1$ and $t \in \mathbf{R}$, and $A_0(t) = \emptyset$. Thus $A_n(t) \in \mathcal{B}_n$ for each $n \geq 1$, and $A_n(t) \subseteq A_{n+1}(t)$. Observe that

(91.26)
$$A_l(t) \setminus A_{l-1}(t) = \{x \in X : f_{l-1}^*(x) \leq t, f_l(x) > t\}$$

when $l \geq 2$, and that

(91.27)
$$A_1(t) \setminus A_0(t) = A_1(t) = \{x \in X : f_1(x) > t\}.$$

In particular, $f_l > t$ on $A_l(t) \setminus A_{l-1}(t)$, and so

(91.28)
$$t \mu(A_l(t) \setminus A_{l-1}(t)) \leq \int_{A_l(t) \setminus A_{l-1}(t)} f_l d\mu.$$

This implies that

(91.29)
$$\begin{aligned} t \mu(A_l(t) \setminus A_{l-1}(t)) &\leq \int_{A_l(t) \setminus A_{l-1}(t)} E(f_n \mid \mathcal{B}_l) d\mu \\ &= \int_{A_l(t) \setminus A_{l-1}(t)} f_n d\mu \end{aligned}$$

when $1 \leq l \leq n$, because $f_l \leq E(f_n \mid \mathcal{B}_l)$, since $\{f_j\}_{j=1}^\infty$ is a submartingale, and $A_l(t) \setminus A_{l-1}(t) \in \mathcal{B}_l$. Of course, the sets $A_l(t) \setminus A_{l-1}(t)$, $1 \leq l \leq n$, are pairwise disjoint, and their union is $A_n(t)$. Hence

(91.30)
$$\begin{aligned} t \mu(A_n(t)) = \sum_{l=1}^n t \mu(A_l(t) \setminus A_{l-1}(t)) &\leq \sum_{l=1}^n \int_{A_l(t) \setminus A_{l-1}(t)} f_n d\mu \\ &= \int_{A_n(t)} f_n d\mu \end{aligned}$$

for each $n \geq 1$ and $t \in \mathbf{R}$.

92 Another variant

Let (X, \mathcal{A}, μ) be a probability space, let $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \cdots$ be an increasing sequence of σ -subalgebras of \mathcal{A} , and let $\{f_j\}_{j=1}^\infty$ be a sequence of functions on X such that $f_j \in L^1(X, \mathcal{B}_j)$ for each j . As in the previous section, put

(92.1)
$$a_j = E(f_{j+1} - f_j \mid \mathcal{B}_j)$$

for each j , $A_l = \sum_{j=1}^{l-1} a_j$ when $l \geq 2$, and $A_1 = 0$. Thus $A_l \in L^1(X, \mathcal{B}_{l-1})$ when $l \geq 2$, and $\{f_l - A_l\}_{l=1}^\infty$ is a martingale, as before. If $f_j \in L^p(X, \mathcal{B}_j)$ for some $p \geq 1$ and each j , and

(92.2)
$$\sum_{j=1}^{\infty} \|f_{j+1} - f_j\|_p$$

converges, then $\{f_j\}_{j=1}^\infty$ converges in the L^p norm and pointwise almost everywhere on X . Suppose instead that $f_j \in L^p(X, \mathcal{B}_j)$ for each j , $\|f_j\|_p$ is bounded, and

$$(92.3) \quad \sum_{j=1}^{\infty} \|a_j\|_p$$

converges. This implies that $\{A_l\}_{l=1}^\infty$ converges in the L^p norm and pointwise almost everywhere on X , and that $\|f_l - A_l\|_p$ is bounded. Because $\{f_l - A_l\}_{l=1}^\infty$ is a martingale, it follows that $\{f_l - A_l\}_{l=1}^\infty$ converges pointwise almost everywhere on X , and in the L^p norm when $1 < p < \infty$.

93 Averaging functions

Let $(X_1, \mathcal{A}_1, \mu_1), (X_2, \mathcal{A}_2, \mu_2), \dots$ be a sequence of probability spaces, and let $X = \prod_{j=1}^\infty X_j$ be their Cartesian product, with the product measure μ . Also let \mathcal{B}_n be the σ -algebra of measurable subsets of X of the form $A \times \prod_{j=n+1}^\infty X_j$, where A is a measurable subset of $\prod_{j=1}^n X_j$. Suppose that $\phi_j \in L^2(X_j, \mathcal{A}_j)$ satisfies

$$(93.1) \quad \int_{X_j} \phi_j d\mu_j = 0$$

and

$$(93.2) \quad \left(\int_{X_j} |\phi_j|^2 d\mu_j \right)^{1/2} \leq C$$

for some $C \geq 0$ and each j , and consider

$$(93.3) \quad f_n(x) = \frac{1}{n} \sum_{j=1}^n \phi_j(x_j),$$

$x = \{x_j\}_{j=1}^\infty \in X$. Thus $f_n \in L^2(X, \mathcal{B}_n)$ for each n , and

$$(93.4) \quad \|f_n\|_2^2 = \frac{1}{n^2} \sum_{j=1}^n \|\phi_j\|_2^2 \leq \frac{C^2}{n},$$

because of orthogonality. In particular, $f_n \rightarrow 0$ in $L^2(X)$ as $n \rightarrow \infty$.

Observe that

$$(93.5) \quad \begin{aligned} f_{n+1}(x) - f_n(x) &= \frac{1}{n+1} \sum_{j=1}^{n+1} \phi_j(x_j) - \frac{1}{n} \sum_{j=1}^n \phi_j(x_j) \\ &= \frac{\phi_{n+1}(x_{n+1})}{n+1} - \frac{1}{n(n+1)} \sum_{j=1}^n \phi_j(x_j). \end{aligned}$$

If $a_n = E(f_{n+1} - f_n \mid \mathcal{B}_n)$, as in the previous section, then

$$(93.6) \quad a_n(x) = -\frac{1}{n(n+1)} \sum_{j=1}^n \phi_j(x_j).$$

This is because $\phi_j(x_j)$ is measurable with respect to \mathcal{B}_n when $j \leq n$, while the conditional expectation of $\phi_{n+1}(x_{n+1})$ with respect to \mathcal{B}_n is equal to 0. Thus $a_n = -(1/(n+1))f_n$,

$$(93.7) \quad \|a_n\|_2 \leq \frac{C}{\sqrt{n(n+1)}},$$

and so $\sum_{n=1}^{\infty} \|a_n\|_2$ converges. It follows that $\{f_n\}_{n=1}^{\infty}$ converges pointwise almost everywhere on X , as in the previous section.

94 Shift mappings

Let $(X_0, \mathcal{A}_0, \mu_0)$ be a probability space, and let X be the space of doubly-infinite sequences $x = \{x_j\}_{j=-\infty}^{\infty}$ with $x_j \in X_0$ for each j . Thus X is the Cartesian product of a family of copies of X_0 indexed by the integers, which is also a probability space with respect to the product measure μ . Let T be the shift mapping on X defined in Section 90, which preserves the measure μ . Also let f be an integrable function on X , and consider

$$(94.1) \quad \frac{f(x) + f(T(x)) + f(T^2(x)) + \cdots + f(T^n(x))}{n+1}.$$

If f is constant, then (94.1) is the same constant for each n . Suppose instead that the integral of f is equal to 0. If f is square-integrable and depends only on one variable, then (94.1) converges to 0 as $n \rightarrow \infty$ in the L^2 norm and pointwise almost everywhere on X , as in the previous section. These are consequences of well-known ergodic theorems as well. One can also deal with other L^p spaces, but let us focus here on $p = 2$ for simplicity. If f depends on only finitely many variables, then one can get the same conclusions from analogous arguments. More precisely, one can begin with averages like (94.1), but using powers of T^r for sufficiently large r in place of powers of T . An average like (94.1) with arbitrary powers of T can then be estimated in terms of r smaller averages involving T^{jr+l} , $l = 0, \dots, r-1$. After that, an arbitrary function f can be approximated by functions depending on only finitely many variables. There are also maximal function estimates for the averages (94.1) like those that have been discussed in other contexts.

95 Families of σ -subalgebras

Let (X, \mathcal{A}, μ) be a probability space, and let (\mathcal{I}, \prec) be a directed system. Thus \mathcal{I} is a set, \prec is a partial ordering on \mathcal{I} , and for each $a, b \in \mathcal{I}$ there is a $c \in \mathcal{I}$ such that $a, b \prec c$. Suppose that for each $a \in \mathcal{I}$ we have a σ -subalgebra \mathcal{B}_a of \mathcal{A} , and that

$$(95.1) \quad \mathcal{B}_a \subseteq \mathcal{B}_b$$

when $a, b \in \mathcal{I}$ and $a \prec b$. If \mathcal{I} is the set \mathbf{Z}_+ of positive integers with the usual ordering, then this is the same as an increasing sequence of σ -subalgebras of \mathcal{A} , as in Section 80.

Alternatively, let I be a nonempty set, and let $(X_i, \mathcal{A}_i, \mu_i)$ be a probability space for each i . Consider the Cartesian product $X = \prod_{i \in I} X_i$ of the X_i 's, with the product measure μ . If \mathcal{I} is the collection of nonempty finite subsets of I , then \mathcal{I} is partially ordered by inclusion, and a directed system. More precisely, if $a, b \in \mathcal{I}$, then $a \cup b \in \mathcal{I}$, and $a, b \subseteq a \cup b$. Let \mathcal{B}_a be the collection of subsets of X that correspond to the Cartesian product of a measurable set $A \subseteq \prod_{i \in a} X_i$ and $\prod_{i \in I \setminus a} X_i$ for each $a \in \mathcal{I}$. It is easy to see that \mathcal{B}_a is a σ -subalgebra of the σ -algebra of measurable subsets of X , and that (95.1) holds. If the X_i 's are compact Hausdorff topological spaces, so that X is also a compact Hausdorff space with respect to the product topology, then one may wish to use Borel sets.

In this product situation, suppose that $\phi_i \in L^1(X_i, \mathcal{A}_i)$ satisfies

$$(95.2) \quad \int_{X_i} \phi_i d\mu_i = 0$$

for each $i \in I$. Put

$$(95.3) \quad \Phi_a(x) = \sum_{i \in a} \phi_i(x_i)$$

for each $a \in \mathcal{I}$, where $x = \{x_i\}_{i \in I} \in X$. Thus $\Phi_a \in L^1(X, \mathcal{B}_a)$, and

$$(95.4) \quad E(\Phi_b \mid \mathcal{B}_a) = \Phi_a$$

when $a, b \in \mathcal{I}$ and $a \subseteq b$. Hence Φ_a , $a \in \mathcal{I}$, defines a martingale with respect to this family of σ -algebras.

Martingales with more general indices like this are discussed in [152]. This point of view is very natural in connection with rearrangement of sums, and convergence of sums in the generalized sense. Note that the arguments for estimating maximal functions as in Section 84 do not work for partially-ordered sets of indices. The corresponding problems with pointwise convergence have already been seen at least implicitly in Section 62, in the case where $I = \mathbf{Z}_+$, $X_i = \{1, -1\}$, and $\mu_i(\{1\}) = \mu_i(\{-1\}) = 1/2$ for each $i \in I$. However, if the σ -algebras \mathcal{B}_a are associated to partitions consisting of intervals in the real line, then one can use a covering argument as in Section 46.

96 Stopping times

Let (X, \mathcal{A}, μ) be a probability space, and let $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \dots$ be an increasing sequence of σ -subalgebras of \mathcal{A} . A function $\tau : X \rightarrow \mathbf{Z}_+$ is said to be a *stopping time* if

$$(96.1) \quad \tau^{-1}(n) = \{x \in X : \tau(x) = n\} \in \mathcal{B}_n$$

for each $n \geq 1$. This is equivalent to the condition that

$$(96.2) \quad \tau^{-1}(\{1, \dots, n\}) = \{x \in X : \tau(x) \leq n\} \in \mathcal{B}_n$$

for each n , since

$$(96.3) \quad \tau^{-1}(\{1, \dots, n\}) = \bigcup_{l=1}^n \tau^{-1}(l)$$

and

$$(96.4) \quad \tau^{-1}(n) = \tau^{-1}(\{1, \dots, n\}) \setminus \tau^{-1}(\{1, \dots, n-1\})$$

when $n \geq 2$. Alternatively, τ is a stopping time if

$$(96.5) \quad \{x \in X : \tau(x) > n\} \in \mathcal{B}_n$$

for each n , because

$$(96.6) \quad \{x \in X : \tau(x) > n\} = X \setminus \tau^{-1}(\{1, \dots, n\}).$$

One can also allow τ to take values in $\mathbf{Z}_+ \cup \{+\infty\}$, in which case

$$(96.7) \quad \tau^{-1}(+\infty) = X \setminus \left(\bigcup_{n=1}^{\infty} \tau^{-1}(n) \right)$$

is in the σ -algebra \mathcal{B}_∞ generated by $\bigcup_{n=1}^{\infty} \mathcal{B}_n$.

If A_1, A_2, \dots is a sequence of pairwise-disjoint subsets of X with $A_n \in \mathcal{B}_n$ for each n , then there is a unique stopping time τ on X such that $\tau^{-1}(n) = A_n$ for each n . More precisely, $\tau(x) < +\infty$ for every $x \in X$ if and only if $\bigcup_{n=1}^{\infty} A_n = X$. Similarly, if $E_1 \subseteq E_2 \subseteq \dots$ is an increasing sequence of subsets of X with $E_n \in \mathcal{B}_n$ for each n , then there is a unique stopping time τ on X such that

$$(96.8) \quad \{x \in X : \tau(x) \leq n\} = E_n$$

for each n . Of course, this corresponds to taking $A_1 = E_1$ and $A_n = E_n \setminus E_{n-1}$ when $n \geq 2$ in the previous statement. As before, $\tau(x) < +\infty$ for every $x \in X$ if and only if $\bigcup_{n=1}^{\infty} E_n = X$.

If τ, τ' are stopping times on X , then $\max(\tau, \tau')$ and $\min(\tau, \tau')$ are stopping times too, because

$$(96.9) \quad \begin{aligned} \{x \in X : \max(\tau(x), \tau'(x)) \leq n\} = \\ \{x \in X : \tau(x) \leq n\} \cap \{x \in X : \tau'(x) \leq n\} \end{aligned}$$

and

$$(96.10) \quad \begin{aligned} \{x \in X : \min(\tau(x), \tau'(x)) \leq n\} = \\ \{x \in X : \tau(x) \leq n\} \cup \{x \in X : \tau'(x) \leq n\}. \end{aligned}$$

In particular,

$$(96.11) \quad \tau_N(x) = \min(\tau(x), N)$$

is a stopping time on X when τ is a stopping time and N is a positive integer.

Suppose that $\{f_n\}_{n=1}^\infty$ is a martingale on X with respect to this filtration, and let

$$(96.12) \quad f^*(x) = \sup_{n \geq 1} |f_n(x)|$$

be the corresponding maximal function. Let $t > 0$ be given, and remember that $f^*(x) > t$ if and only if $|f_n(x)| > t$ for some n . If $f^*(x) > t$, then let $\tau(x)$ be the smallest positive integer such that

$$(96.13) \quad |f_{\tau(x)}(x)| > t,$$

and put $\tau(x) = +\infty$ when $f^*(x) \leq t$. Thus $\tau(x) = n$ exactly when $|f_n(x)| > t$ and $|f_l(x)| \leq t$ for $l < n$. This implies that $\tau^{-1}(n) \in \mathcal{B}_n$ for each n , because f_l is measurable with respect to $\mathcal{B}_l \subseteq \mathcal{B}_n$ when $l \leq n$.

Let τ be a stopping time on X such that $\tau(x) < +\infty$ for every $x \in X$, and let \mathcal{B}_τ be the collection of subsets A of X such that

$$(96.14) \quad A \cap \tau^{-1}(n) \in \mathcal{B}_n$$

for each n . It is easy to see that this is a σ -algebra, because \mathcal{B}_n is a σ -algebra for each n , and that $\mathcal{B}_\tau \subseteq \mathcal{B}_\infty$. If N is a positive integer and $\tau(x) \leq N$ for each $x \in X$, then

$$(96.15) \quad \mathcal{B}_\tau \subseteq \mathcal{B}_N.$$

More precisely, if τ is any finite stopping time, $A \in \mathcal{B}_\tau$, and $\tau(x) \leq N$ for every $x \in A$, then $A \in \mathcal{B}_N$. If τ' is another stopping time such that

$$(96.16) \quad \tau(x) \leq \tau'(x) < +\infty$$

for every $x \in X$, then $\mathcal{B}_\tau \subseteq \mathcal{B}_{\tau'}$.

Let $\{f_n\}_{n=1}^\infty$ be a martingale on X with respect to this filtration, and let τ be a finite stopping time on X . If f_τ is the function on X defined by

$$(96.17) \quad f_\tau(x) = f_{\tau(x)}(x),$$

then f_τ is measurable with respect to \mathcal{B}_τ , because f_n is measurable with respect to \mathcal{B}_n for each n . Let us check that

$$(96.18) \quad \int_{\tau^{-1}(\{1, \dots, N\})} |f_\tau| d\mu \leq \int_X |f_N| d\mu$$

for each positive integer N . By the definition of f_τ ,

$$(96.19) \quad \int_{\tau^{-1}(\{1, \dots, N\})} |f_\tau| d\mu = \sum_{n=1}^N \int_{\tau^{-1}(n)} |f_n| d\mu.$$

Hence

$$(96.20) \quad \begin{aligned} \int_{\tau^{-1}(\{1, \dots, N\})} |f_\tau| d\mu &\leq \sum_{n=1}^N \int_{\tau^{-1}(n)} |f_N| d\mu \\ &= \int_{\tau^{-1}(\{1, \dots, N\})} |f_N| d\mu, \end{aligned}$$

because $|f_n| \leq E(|f_N| \mid \mathcal{B}_n)$ when $n \leq N$.

If $\tau(x) \leq N$ for every $x \in X$, then it follows that f_τ is integrable on X . Let us verify that

$$(96.21) \quad f_\tau = E(f_N \mid \mathcal{B}_\tau),$$

remembering that $\mathcal{B}_\tau \subseteq \mathcal{B}_N$ in this case. To see this, it suffices to show that

$$(96.22) \quad \int_A f_\tau d\mu = \int_A f_N d\mu$$

for every $A \in \mathcal{B}_\tau$. Under these conditions,

$$(96.23) \quad \begin{aligned} \int_A f_\tau d\mu &= \sum_{n=1}^N \int_{A \cap \tau^{-1}(n)} f_n d\mu \\ &= \sum_{n=1}^N \int_{A \cap \tau^{-1}(n)} f_N d\mu = \int_A f_N d\mu, \end{aligned}$$

because $A \cap \tau^{-1}(n) \in \mathcal{B}_n$ and $f_n = E(f_N \mid \mathcal{B}_n)$ when $n \leq N$.

Similarly, if the f_n 's have bounded L^1 norms and τ is any finite stopping time on X , then we get that

$$(96.24) \quad \int_{\tau^{-1}(\{1, \dots, N\})} |f_\tau| d\mu \leq \sup_{n \geq 1} \int_X |f_n| d\mu$$

for every positive integer N . This implies that f_τ is integrable on X , and that

$$(96.25) \quad \int_X |f_\tau| d\mu \leq \sup_{n \geq 1} \int_X |f_n| d\mu.$$

In particular, this holds when there is an $f \in L^1(X, \mathcal{A})$ such that $f_n = E(f \mid \mathcal{B}_n)$ for each n . In this case, one can check that

$$(96.26) \quad f_\tau = E(f \mid \mathcal{B}_\tau).$$

As before, one can show that

$$(96.27) \quad \int_A f_\tau d\mu = \int_A f d\mu$$

when $A \in \mathcal{B}_\tau$, by expressing A as the union of $A \cap \tau^{-1}(n)$, $n \geq 1$, and using the fact that $f_\tau = f_n = E(f \mid \mathcal{B}_n)$ on $A \cap \tau^{-1}(n) \in \mathcal{B}_n$.

Now let τ be a stopping time on X that takes values in $\mathbf{Z}_+ \cup \{+\infty\}$, so that $\tau_N = \min(\tau, N)$ is a finite stopping time on X for each N . Let $\{f_n\}_{n=1}^\infty$ be a martingale on X with respect to this filtration, and note that f_{τ_N} is integrable on X for each N , since τ_N is bounded. Let us check that

$$(96.28) \quad f_{\tau_N} = E(f_{\tau_{N+1}} \mid \mathcal{B}_N)$$

for each N , so that $\{f_{\tau_N}\}_{N=1}^\infty$ is a martingale as well. As usual, we would like to show that

$$(96.29) \quad \int_A f_{\tau_N} d\mu = \int_A f_{\tau_{N+1}} d\mu$$

when $A \in \mathcal{B}_N$. Consider

$$(96.30) \quad A_1 = \{x \in A : \tau(x) \leq N\}$$

and

$$(96.31) \quad A_2 = \{x \in A : \tau(x) > N\}.$$

Thus $A_1 \cup A_2 = A$, $A_1 \cap A_2 = \emptyset$, and

$$(96.32) \quad A_1, A_2 \in \mathcal{B}_N,$$

since τ is a stopping time. If $x \in A_1$, then $\tau_N(x) = \tau_{N+1}(x) = \tau(x)$, and hence $f_{\tau_N}(x) = f_{\tau_{N+1}}(x) = f_\tau(x)$. This implies that

$$(96.33) \quad \int_{A_1} f_{\tau_N} d\mu = \int_{A_1} f_{\tau_{N+1}} d\mu.$$

If $x \in A_2$, then $\tau_N(x) = N$, $\tau_{N+1}(x) = N+1$, and so $f_{\tau_N}(x) = f_N(x)$, $f_{\tau_{N+1}}(x) = f_{N+1}(x)$. It follows that

$$(96.34) \quad \int_{A_2} f_{\tau_N} d\mu = \int_{A_2} f_N d\mu = \int_{A_2} f_{N+1} d\mu = \int_{A_2} f_{\tau_{N+1}} d\mu,$$

because $A_2 \in \mathcal{B}_N$ and $f_N = E(f_{N+1} \mid \mathcal{B}_N)$, as desired.

If $\tau(x) < \infty$ for every $x \in X$, then $\{f_{\tau_N}\}_{N=1}^\infty$ converges to f_τ pointwise on X , because $f_{\tau_N}(x) = f_\tau(x)$ when $N \geq \tau(x)$. If the f_n 's have bounded L^1 norms, then the f_{τ_N} 's also have bounded L^1 norms, and f_τ is integrable. A necessary and sufficient condition for $\{f_{\tau_N}\}_{N=1}^\infty$ to converge to f_τ in the L^1 norm is that

$$(96.35) \quad \int_{\{x \in X : \tau(x) > N\}} |f_{\tau_N}(x)| d\mu(x) = \int_{\{x \in X : \tau(x) > N\}} |f_N(x)| d\mu(x) \rightarrow 0$$

as $N \rightarrow \infty$. This holds automatically when the f_n 's are uniformly integrable, and otherwise depends on both the f_n 's and τ .

97 Ultrametrics

A metric $d(x, y)$ on a set M is said to be an *ultrametric* if

$$(97.1) \quad d(x, z) \leq \max(d(x, y), d(y, z))$$

for every $x, y, z \in M$. If X_1, X_2, \dots is a sequence of nonempty sets, and r_1, r_2, \dots is a decreasing sequence of positive real numbers that converges to 0, then one can define an ultrametric on the Cartesian product $X = \prod_{j=1}^\infty X_j$ as follows.

Each element x of X is a sequence $\{x_j\}_{j=1}^{\infty}$ with $x_j \in X_j$ for every j , and we put $d(x, x) = 0$, and

$$(97.2) \quad d(x, y) = r_l$$

when $x \neq y$ and l is the smallest positive integer such that $x_l \neq y_l$. It is easy to see that this is an ultrametric on X , and that the corresponding topology is the product topology associated to the discrete topology on X_j for each j .

If $d(x, y)$ is an ultrametric on a set M , $p, q \in M$, and $t \geq r > 0$, then either

$$(97.3) \quad B(p, r) \subseteq B(q, t) \quad \text{or} \quad B(p, r) \cap B(q, t) = \emptyset.$$

More precisely, the first alternative holds when $d(p, q) < t$, and the second alternative holds when $d(p, q) \geq t$. Using this, one can check that open balls are closed subsets of ultrametric spaces. There is an analogous dichotomy for closed balls, which implies that closed balls are open subsets of ultrametric spaces. It follows that ultrametric spaces are totally disconnected, in the sense that they do not contain connected subsets with more than one element.

Another consequence of the previous dichotomy is that

$$(97.4) \quad B(p, r) = B(q, r)$$

when $d(p, q) < r$. Thus every element of an open ball in M can be used as a center of that ball. The collection of open balls in M with the same radius r forms a partition of M , because any two such balls are either the same or disjoint as subsets of M . If $t \geq r$, then the partition of M into open balls of radius r is a refinement of the partition of M into open balls of radius t , since every ball of radius r is contained in a ball of radius t .

The geometry of an ultrametric space is very similar to a probability space with an increasing sequence of σ -subalgebras of the σ -algebra of measurable sets. In particular, one can consider σ -subalgebras of the Borel sets in an ultrametric space corresponding to partitions by balls of a given radius. One can also deal directly with Hardy–Littlewood type maximal functions, using the nesting properties of balls to reduce of covering of a set by balls of bounded radius to a disjoint union of balls that are maximal elements of the covering. Of course, there are more complicated covering arguments for Euclidean spaces and other metric spaces, including the basic property of intervals in the real line mentioned in Section 46. These can also be used to estimate maximal functions, and so on.

Part IV

Vector-valued functions

98 Some randomized sums

Let (X, \mathcal{A}, μ) be a probability space, and let ϕ_1, \dots, ϕ_n be bounded real or complex-valued measurable functions on X , with

$$(98.1) \quad \|\phi_j\|_\infty \leq C$$

for some $C \geq 0$ and $j = 1, \dots, n$. Also let $\{1, -1\}^n$ be the set of sequences $\epsilon = \{\epsilon_j\}_{j=1}^n$ of length n with $\epsilon_j = 1$ or -1 for each j . If $2 \leq p < \infty$, then there is a positive real number $C(p)$ such that

$$(98.2) \quad 2^{-n} \sum_{\epsilon \in \{1, -1\}^n} \int_X \left| \sum_{j=1}^n \epsilon_j a_j \phi_j(x) \right|^p d\mu(x) \leq C(p) \left(\sum_{j=1}^n |a_j|^2 \right)^{p/2}$$

for all $a_1, \dots, a_n \in \mathbf{R}$ or \mathbf{C} , as appropriate. Of course, the left side is the same as

$$(98.3) \quad \int_X 2^{-n} \sum_{\epsilon \in \{1, -1\}^n} \left| \sum_{j=1}^n \epsilon_j a_j \phi_j(x) \right|^p d\mu(x).$$

As in Section 61,

$$(98.4) \quad 2^{-n} \sum_{\epsilon \in \{1, -1\}^n} \left| \sum_{j=1}^n \epsilon_j a_j \phi_j(x) \right|^p \leq C'(p) \left(\sum_{j=1}^n |a_j \phi_j(x)|^2 \right)^{p/2}$$

for some $C'(p) > 0$ and all $a_1, \dots, a_n \in \mathbf{R}$ or \mathbf{C} and $x \in X$. This implies (98.2), by integrating in x and using the uniform boundedness of the ϕ_j 's. More precisely, $C(p)$ depends only on C and p , and not on a_1, \dots, a_n or n .

If $p = 2$, then we have that

$$(98.5) \quad 2^{-n} \sum_{\epsilon \in \{1, -1\}^n} \left| \sum_{j=1}^n \epsilon_j a_j \phi_j(x) \right|^2 = \sum_{j=1}^n |a_j \phi_j(x)|^2.$$

This implies that

$$(98.6) \quad 2^{-n} \sum_{\epsilon \in \{1, -1\}^n} \int_X \left| \sum_{j=1}^n \epsilon_j a_j \phi_j \right|^2 d\mu(x) = \sum_{j=1}^n |a_j|^2$$

when $\|\phi_j\|_2 = 1$ for each j . Otherwise, if $\|\phi_j\|_2 \geq c$ for some $c > 0$ and each j , then we get that

$$(98.7) \quad 2^{-n} \sum_{\epsilon \in \{1, -1\}^n} \int_X \left| \sum_{j=1}^n \epsilon_j a_j \phi_j \right|^2 d\mu(x) \geq c^2 \sum_{j=1}^n |a_j|^2.$$

Note that

$$(98.8) \quad \left(2^{-n} \sum_{\epsilon \in \{1, -1\}} \int_X \left| \sum_{j=1}^n \epsilon_j a_j \phi_j(x) \right|^p d\mu(x) \right)^{1/p}$$

is monotone increasing in p , by Jensen's inequality. This is the same as the L^p norm of $\sum_{j=1}^n \epsilon_j a_j \phi_j(x)$ as a function of (x, ϵ) on $X \times \{1, -1\}^n$, with respect to the product of μ on X and 2^{-n} times counting measure on $\{1, -1\}^n$.

Under these conditions, if $0 < p < 2$, then there is a $C(p) > 0$ such that

$$(98.9) \quad C(p)^{-1} \left(\sum_{j=1}^n |a_j|^2 \right)^{p/2} \leq 2^{-n} \sum_{\epsilon \in \{1, -1\}^n} \int_X \left| \sum_{j=1}^n \epsilon_j a_j \phi_j \right|^p d\mu(x)$$

for all $a_1, \dots, a_n \in \mathbf{R}$ or \mathbf{C} . This can be derived from the previous estimates and Hölder's inequality, as in Section 61. More precisely, Hölder's inequality can be used to estimate the L^2 norm of $\sum_{j=1}^n \epsilon_j a_j \phi_j(x)$ on $X \times \{1, -1\}^n$ in terms of its L^p and L^4 norms, as before. Under the present conditions, the L^2 norm is bounded from below by a constant multiple of $\left(\sum_{j=1}^n |a_j|^2 \right)^{1/2}$, and the L^4 norm is bounded from above by a multiple of the same expression, which leads to a lower bound for the L^p norm as in (98.9). As usual, the constant $C(p)$ in (98.9) depends on c , C , and p , and not on a_1, \dots, a_n or n .

99 Randomized sums, 2

Let (X, \mathcal{A}, μ) be a probability space again, and let ϕ_1, \dots, ϕ_n be orthonormal functions in $L^2(X)$. As usual, this implies that

$$(99.1) \quad \int_X \left| \sum_{j=1}^n \alpha_j \phi_j(x) \right|^2 d\mu(x) = \sum_{j=1}^n |\alpha_j|^2$$

for all $\alpha_1, \dots, \alpha_n \in \mathbf{R}$ or \mathbf{C} , as appropriate. Hence

$$(99.2) \quad \int_X \left| \sum_{j=1}^n \epsilon_j \alpha_j \phi_j(x) \right|^2 d\mu(x) = \sum_{j=1}^n |\alpha_j|^2$$

for every $\epsilon \in \{1, -1\}^n$. In particular, the average of the left side of (99.2) over $\epsilon \in \{1, -1\}^n$ has the same value, as in (98.6).

Suppose that ϕ_1, ϕ_2, \dots is an orthonormal basis for $L^2(X)$, and that the ϕ_j 's are uniformly bounded on X , as in the previous section. Thus every function in $L^2(X)$ can be approximated in the L^2 norm by a finite sum of the form

$$(99.3) \quad \sum_{j=1}^n \alpha_j \phi_j(x).$$

Moreover, the average of the L^p norms of

$$(99.4) \quad \sum_{j=1}^n \epsilon_j \alpha_j \phi_j(x)$$

over $\epsilon \in \{1, -1\}^n$ is bounded by a constant multiple of the L^2 norm for every $p < \infty$, as before. However, this does not mean that the L^p norm of (99.4) is bounded by a multiple of the L^2 norm for every $\epsilon \in \{1, -1\}$, or even for only $\epsilon = (1, \dots, 1)$. If we start with a function in $L^2(X)$ which is not in $L^p(X)$ for some $p > 2$, then the L^p norms of its approximations are necessarily unbounded. Note that Fourier series and Walsh functions are examples of this type of situation. Lacunary series and Rademacher functions correspond to subsets of these bases for which the L^p norms are bounded by constant multiples of the L^2 norms when $2 < p < \infty$.

100 The unit square

Let $X = [0, 1) \times [0, 1)$ be the version of the unit square associated to dyadic intervals, equipped with 2-dimensional Lebesgue measure. If $I, L \subseteq [0, 1)$ are dyadic intervals with the same length 2^{-j} , then their Cartesian product $I \times L$ is a dyadic square in X with side length 2^{-j} and area 2^{-2j} . There are 2^{2j} dyadic squares in X with side length 2^{-j} , they are pairwise disjoint, and their union is equal to X . Let \mathcal{A}_j be the collection of subsets of X which can be expressed as unions of dyadic squares with side length 2^{-j} , including the empty set. This is the same as the σ -algebra of subsets of X generated by the partition \mathcal{P}_j of X into dyadic squares of side length 2^{-j} , as in Section 77. Note that \mathcal{A}_j is a σ -subalgebra of the σ algebra of Borel subsets of X , and that $\mathcal{A}_j \subseteq \mathcal{A}_{j+1}$ for each j . As usual, a function on X is measurable with respect to \mathcal{A}_j if and only if it is constant on dyadic squares with side length 2^{-j} .

Let $f_j(x, y)$ be the function on X defined by

$$(100.1) \quad f_j(x, y) = 2^j$$

when x, y are contained in the same dyadic interval of length 2^{-j} , and

$$(100.2) \quad f_j(x, y) = 0$$

when x, y are contained in distinct dyadic intervals of length 2^{-j} . In particular,

$$(100.3) \quad \int_{I \times I} f_j(x, y) dx dy = 2^{-j}$$

for each dyadic interval I of length 2^{-j} . Summing over I , we get that

$$(100.4) \quad \int_{[0,1) \times [0,1)} f_j(x, y) dx dy = 1$$

for each j , because there are 2^j dyadic intervals of length 2^{-j} . Clearly $f_j(x, y)$ is measurable with respect to \mathcal{A}_j for each j . It is easy to see that

$$(100.5) \quad f_j = E(f_{j+1} \mid \mathcal{A}_j)$$

for each j , so that $\{f_j\}_j$ is a martingale with respect to the \mathcal{A}_j 's.

Let ν be the Borel measure on X defined by

$$(100.6) \quad \nu(A) = |\{x \in [0, 1) : (x, x) \in A\}|,$$

where $|E|$ denotes the Lebesgue measure of $E \subseteq [0, 1)$. Alternatively, if

$$(100.7) \quad \Delta = \{(x, x) : x \in [0, 1)\}$$

is the diagonal in X , then

$$(100.8) \quad \nu(A) = |\pi(A \cap \Delta)|,$$

where $\pi(x, x) = x$ is the natural projection of Δ onto $[0, 1)$. Of course, the restriction of ν to \mathcal{A}_j is absolutely continuous with respect to the restriction of 2-dimensional Lebesgue measure to \mathcal{A}_j for each j . One can also think of f_j as the conditional expectation of ν with respect to \mathcal{A}_j , as in Section 86.

If $x, y \in [0, 1)$ and $x \neq y$, then $f_j(x, y) = 0$ for all sufficiently large j . In particular, $\{f_j(x, y)\}_j$ converges to 0 almost everywhere on X . Basically, $\{f_j\}_j$ converges to ν in a suitable weak sense.

Now let \mathcal{B}_j be the collection of subsets of X that can be expressed as the union of sets of the form $I \times A(I)$, where I runs through the dyadic subintervals of $[0, 1)$ of length 2^{-j} , and $A(I)$ is a Borel set in $[0, 1)$ for each such I . Equivalently, $A \in \mathcal{B}_j$ if for each dyadic interval $I \subseteq [0, 1)$ with $|I| = 2^{-j}$ there is a Borel set $A(I) \subseteq [0, 1)$ such that

$$(100.9) \quad A \cap (I \times [0, 1)) = I \times A(I).$$

Thus \mathcal{B}_j is a σ -subalgebra of the σ -algebra of Borel sets in X , $\mathcal{A}_j \subseteq \mathcal{B}_j$, and $\mathcal{B}_j \subseteq \mathcal{B}_{j+1}$ for each j . A function $f(x, y)$ on X is measurable with respect to \mathcal{B}_j if and only if it is constant in x on each dyadic interval I of length 2^{-j} and Borel measurable in y .

In particular, $f_j(x, y)$ is measurable with respect to \mathcal{B}_j for each j . One can also check that

$$(100.10) \quad E(f_{j+1} \mid \mathcal{B}_j) = f_j$$

for each j , so that $\{f_j\}_j$ is a martingale with respect to the \mathcal{B}_j 's as well. The main point is that

$$(100.11) \quad \int_{I \times A} f_j(x, y) dx dy = |A \cap I|$$

for each dyadic interval I of length 2^{-j} and Borel set $A \subseteq [0, 1)$. Similarly, if I_1, I_2 are the dyadic intervals of length 2^{-j-1} such that $I = I_1 \cup I_2$, then

$$(100.12) \quad \int_{I \times A} f_{j+1}(x, y) dx dy$$

$$\begin{aligned}
&= \int_{I_1 \times A} f_{j+1}(x, y) dx dy + \int_{I_2 \times A} f_{j+1}(x, y) dx dy \\
&= |A \cap I_1| + |A \cap I_2| = |A \cap I|.
\end{aligned}$$

This implies (100.10), which can also be seen by viewing f_j as the conditional expectation of ν with respect to \mathcal{B}_j , by (100.11).

Let Φ_j be the function on $[0, 1)$ with values in $L^1([0, 1))$ defined by

$$(100.13) \quad \Phi_j(x)(y) = f_j(x, y).$$

This may be considered as a martingale on $[0, 1)$ with values in $L^1([0, 1))$, with respect to the usual filtration associated to dyadic intervals of length 2^{-j} . Note that the L^1 norm of $\Phi_j(x)$ is equal to 1 for each x and j , but $\{\Phi_j(x)\}_j$ does not converge in $L^1([0, 1))$ for any $x \in [0, 1)$. If we identify integrable functions on $[0, 1)$ with absolutely continuous Borel measures on $[0, 1]$, which determine bounded linear functionals on the space of continuous functions on $[0, 1]$ with respect to the supremum norm, then $\{\Phi_j(x)\}_j$ converges in the weak* topology to the Dirac mass at x .

101 Partitions and products

Let $(X_1, \mathcal{A}_1, \mu_1)$, $(X_2, \mathcal{A}_2, \mu_2)$ be probability spaces, and let $X = X_1 \times X_2$ be their Cartesian product, with the product probability measure $\mu = \mu_1 \times \mu_2$. Suppose that \mathcal{P}_1 , \mathcal{P}_2 are partitions of X_1 , X_2 into finitely or countably many measurable sets, respectively, as in Section 77. The corresponding product partition $\mathcal{P}_{1,2}$ of X consists of all products $A_1 \times A_2$, with $A_1 \in \mathcal{P}_1$ and $A_2 \in \mathcal{P}_2$. It is easy to see that this is a partition of X into finitely or countably many measurable sets, and that the σ -algebra generated by $\mathcal{P}_{1,2}$ is the same as the one associated to the σ -algebras generated by \mathcal{P}_1 , \mathcal{P}_2 in the product space. A function $f(x_1, x_2)$ on X is measurable with respect to this σ -algebra if and only if it is constant on $A_1 \times A_2$ for each $A_1 \in \mathcal{P}_1$ and $A_2 \in \mathcal{P}_2$.

Now let \mathcal{P}_1 be a partition of X_1 into finitely or countably many measurable sets, and let \mathcal{B}_2 be a σ -subalgebra of \mathcal{A}_2 . This leads to a σ -subalgebra $\mathcal{B}_{1,2}$ of the σ -algebra of measurable subsets of X associated to the σ -algebra generated by \mathcal{P}_1 and \mathcal{B}_2 in the product space. As in the special case described in the previous section, $\mathcal{B}_{1,2}$ consists of the sets $A \subseteq X$ such that for each $A_1 \in \mathcal{P}_1$ there is an $A_2 \in \mathcal{B}_2$ such that

$$(101.1) \quad A \cap (A_1 \times X_2) = A_1 \times A_2.$$

Equivalently, $A \in \mathcal{B}_{1,2}$ if A can be expressed as a union of sets of the form $A_1 \times A_2$, where A_1 runs through the elements of \mathcal{P}_1 , and $A_2 \in \mathcal{B}_2$ for each $A_1 \in \mathcal{P}_1$. Thus a function $f(x_1, x_2)$ on X is measurable with respect to $\mathcal{B}_{1,2}$ if it is constant in x_1 on each $A_1 \in \mathcal{P}_1$, and measurable in x_2 with respect to \mathcal{B}_2 for each $x_1 \in X_1$.

As in Section 77, it will be convenient to ask that $\mu_1(A_1) > 0$ for each $A_1 \in \mathcal{P}_1$. If $\mathcal{B}_2 = \mathcal{A}_2$ and f is an integrable function on X , then the conditional expectation of f with respect to $\mathcal{B}_{1,2}$ is given by

$$(101.2) \quad E(f \mid \mathcal{B}_{1,2})(x_1, x_2) = \frac{1}{\mu(A_1)} \int_{A_1} f(t, x_2) d\mu_1(t)$$

when $x_1 \in A_1 \in \mathcal{P}_1$. This can be seen as a combination of the conditional expectations associated to partitions and product spaces, as in Sections 75 and 77. If \mathcal{B}_2 is any σ -subalgebra of \mathcal{A}_2 , then $E(f \mid \mathcal{B}_{1,2})$ can be obtained by first averaging $f(x_1, x_2)$ over $x_1 \in A_1$ for each $A_1 \in \mathcal{P}_1$, as before, and then taking the conditional expectation of the resulting functions of x_2 with respect to \mathcal{B}_2 . In this case, $\mathcal{B}_{1,2}$ is a σ -subalgebra of the σ -algebra associated to \mathcal{P}_1 and \mathcal{A}_2 .

102 Partitions and vectors

Let (X, \mathcal{A}, μ) be a probability space, and let \mathcal{P} be a partition of X into finitely or countably many measurable sets, as in Section 77. As usual, it will be convenient to ask that $\mu(A) > 0$ for each $A \in \mathcal{P}$. Also let $\mathcal{B}(\mathcal{P})$ be the σ -subalgebra of \mathcal{A} generated by \mathcal{P} , consisting of unions of elements of \mathcal{P} , including the empty set. Thus a function on X is measurable with respect to $\mathcal{B}(\mathcal{P})$ if and only if it is constant on the elements of \mathcal{P} .

Let V be a real or complex vector space with a norm $\|v\|$, and let $f(x)$ be a V -valued function on X that is constant on the elements of \mathcal{P} . In particular, $\|f(x)\|$ is a nonnegative real-valued function on X that is constant on the elements of \mathcal{P} . If $f(A)$ denotes the value of f on $A \in \mathcal{P}$, then

$$(102.1) \quad \int_X \|f(x)\| d\mu(x) = \sum_{A \in \mathcal{P}} \|f(A)\| \mu(A).$$

More precisely, if \mathcal{P} is a partition of X into finitely many sets, then the sum on the right is a finite sum, and $\|f(x)\|$ is automatically integrable on X . If \mathcal{P} consists of infinitely many measurable subsets of X , then the sum on the right is interpreted as the supremum of the corresponding sums over finite subsets of \mathcal{P} , which may be infinite.

If \mathcal{P} has only finitely many elements, then we can put

$$(102.2) \quad \int_X f(x) d\mu(x) = \sum_{A \in \mathcal{P}} f(A) \mu(A).$$

This also makes sense when \mathcal{P} has infinitely many elements, $\|f(x)\|$ is integrable on X , and V is complete. In this case, the sum on the right side of (102.1) is finite, and the sum on the right side of (102.2) converges in the generalized sense, as in Section 15. In both cases,

$$(102.3) \quad \left\| \int_X f(x) d\mu(x) \right\| \leq \int_X \|f(x)\| d\mu(x).$$

Similarly, if $B \in \mathcal{B}(\mathcal{P})$, then we would like to put

$$(102.4) \quad \int_B f(x) d\mu(x) = \sum_{\substack{A \in \mathcal{P} \\ A \subseteq B}} f(A) \mu(A).$$

As before, this makes sense when B is the union of finitely many elements of \mathcal{P} , and when B contains infinitely many elements of \mathcal{P} , $\|f(x)\|$ is integrable, and V is complete. We also have the analogue of (102.3) with $X = B$.

Using the Bochner integral, one can integrate much more complicated vector-valued functions. We shall restrict our attention here to sums over partitions for the sake of simplicity.

103 Vector-valued martingales

Let (X, \mathcal{A}, μ) be a probability space, and suppose that $\mathcal{P}_1, \mathcal{P}_2, \dots$ is a sequence of partitions of X into finitely or countably many measurable subsets such that \mathcal{P}_{j+1} is a refinement of \mathcal{P}_j for each j . This means that each $B \in \mathcal{P}_j$ is the union of the $A \in \mathcal{P}_{j+1}$ such that $A \subseteq B$. If $\mathcal{B}_j = \mathcal{B}(\mathcal{P}_j)$ is the σ -algebra generated by \mathcal{P}_j , then it follows that $\mathcal{B}_j \subseteq \mathcal{B}_{j+1}$ for each j . As usual, it is convenient to ask that $\mu(A) > 0$ for each $A \in \mathcal{P}_j$.

Let V be a real or complex vector space with a norm $\|v\|$, and let f_l is a V -valued function on X that is constant on elements of \mathcal{P}_l . We would like to define the conditional expectation of f_l with respect to \mathcal{B}_j for $j < l$ by

$$(103.1) \quad E(f_l | \mathcal{B}_j)(x) = \frac{1}{\mu(B)} \int_B f_l d\mu = \sum_{\substack{A \in \mathcal{P}_l \\ A \subseteq B}} f_l(A) \frac{\mu(A)}{\mu(B)}$$

when $x \in B \in \mathcal{P}_j$, where $f_l(A)$ denotes the value of f_l on $A \in \mathcal{P}_l$, as in the previous section. This makes sense when each $B \in \mathcal{P}_j$ is the union of finitely many $A \in \mathcal{P}_l$, and when $\|f_l\|$ is integrable and V is complete. In both cases, it is easy to see that

$$(103.2) \quad \|E(f_l | \mathcal{B}_j)\| \leq E(\|f_l\| | \mathcal{B}_j).$$

If $j < k < l$, then one can also check that

$$(103.3) \quad E(E(f_l | \mathcal{B}_k) | \mathcal{B}_j) = E(f_l | \mathcal{B}_j),$$

under these conditions, just as in the context of real or complex-valued functions.

Now let $\{f_j\}_{j=1}^\infty$ be a sequence of V -valued functions on X such that f_j is constant on the elements of \mathcal{P}_j for each j . As usual, $\{f_j\}_{j=1}^\infty$ is said to be a martingale with respect to this filtration if

$$(103.4) \quad f_j = E(f_l | \mathcal{B}_j)$$

for each $j \leq l$. More precisely, this makes sense when each element of \mathcal{P}_j is the union of finitely many elements of \mathcal{P}_l , and when each $\|f_l\|$ is integrable and V

is complete. Note that (103.4) holds for all $j \leq l$ when it holds for $l = j + 1$, because of (103.3).

Of course, the simplest type of situation occurs when \mathcal{P}_j consists of only finitely many measurable subsets of X for each j . All of the sums involved in the conditional expectations are then finite sums, and the functions $\|f_l\|$ are automatically bounded.

104 L^1 -Valued martingales

Let us continue with the same notations and hypotheses as in the previous section. As in Section 100, we can get an example of a V -valued martingale on X with $V = L^1(X, \mathcal{A})$ by taking

$$(104.1) \quad f_l(A) = \mu(A)^{-1} \mathbf{1}_A$$

for each $A \in \mathcal{P}_l$. Here $\mathbf{1}_A$ denotes the indicator function associated to A on X , equal to 1 on A and 0 on $X \setminus A$, as usual. Thus $\|f_l(x)\|_1 = 1$ for every $x \in X$ and $l \geq 1$, and it is easy to check that (103.4) holds.

Now let (Y, \mathcal{B}, ν) be a σ -finite measure space, and let us consider functions on X with values in $V = L^1(Y)$. If $f_l(x)$ is an $L^1(Y)$ -valued function on X that is constant on the elements of \mathcal{P}_l , then

$$(104.2) \quad F_l(x, y) = f_l(x)(y)$$

defines a function on $X \times Y$ that is constant in x on each element of \mathcal{P}_l and measurable in y for each $x \in X$. If $\|f_l(x)\|_{L^1(Y)}$ is integrable on X , then $F_l(x, y)$ is integrable on $X \times Y$, and

$$(104.3) \quad \begin{aligned} \int_X \|f_l(x)\|_{L^1(Y)} d\mu(x) &= \int_X \left(\int_Y |F_l(x, y)| d\nu(y) \right) d\mu(x) \\ &= \int_{X \times Y} |F_l(x, y)| d(\mu \times \nu)(x, y). \end{aligned}$$

Conversely, if $F_l(x, y)$ is an integrable function on $X \times Y$ that is constant in x on each element of \mathcal{P}_l , then we get an $L^1(Y)$ -valued function $f_l(x)$ on X that is constant on each element of \mathcal{P}_l and for which $\|f_l(x)\|_{L^1(Y)}$ is integrable on X .

Let $\widehat{\mathcal{B}}_l$ be the σ -algebra of subsets of $X \times Y$ that corresponds to $\mathcal{B}_l = \mathcal{B}(\mathcal{P}_l)$ on X and \mathcal{B} on Y in the product space. As in Section 101, a set $\widehat{A} \subseteq X \times Y$ is in $\widehat{\mathcal{B}}_l$ if and only if for each $A \in \mathcal{P}_l$ there is a $B \in \mathcal{B}$ such that

$$(104.4) \quad \widehat{A} \cap (A \times Y) = A \times B.$$

Equivalently, $\widehat{A} \in \widehat{\mathcal{B}}_l$ if it can be expressed as the union of sets of the form $A \times B(A)$, where A runs through the elements of \mathcal{P}_l , and $B(A) \in \mathcal{B}$ for each $A \in \mathcal{P}_l$. In the context of the preceding paragraph, the functions $F_l(x, y)$ are measurable with respect to $\widehat{\mathcal{B}}_l$.

Suppose that $\{f_l\}_{l=1}^\infty$ is a sequence of $L^1(Y)$ -valued functions on X such that $f_l(x)$ is constant on each element of \mathcal{P}_l and $\|f_l(x)\|_{L^1(Y)}$ is integrable on X for each l . This corresponds exactly to a sequence $\{F_l\}_{l=1}^\infty$ of integrable functions on $X \times Y$ such that $F_l(x, y)$ is measurable with respect to $\widehat{\mathcal{B}}_l$ for each l , as in the previous paragraphs. If Y is a probability space, then $X \times Y$ is also a probability space, and it is easy to see that $\{f_l\}_{l=1}^\infty$ is an $L^1(Y)$ -valued martingale on X with respect to the \mathcal{B}_l 's if and only if $\{F_l\}_{l=1}^\infty$ is a martingale on $X \times Y$ with respect to the $\widehat{\mathcal{B}}_l$'s. This basically works as well when Y is σ -finite, by extending the relevant definitions in a natural way.

105 Pointwise convergence

Let us continue with the same notation and hypotheses as in Section 103, with the additional condition that V be complete. Suppose that $\{f_j\}_{j=1}^\infty$ is a sequence of V -valued functions on X such that $f_j(x)$ is constant on each element of \mathcal{P}_j , $\|f_j(x)\|$ is integrable on X for each j , and $\{f_j\}_{j=1}^\infty$ is a martingale with respect to $\mathcal{B}_j = \mathcal{B}(\mathcal{P}_j)$. If

$$(105.1) \quad f_n^*(x) = \max_{1 \leq j \leq n} \|f_j(x)\|$$

is the usual maximal function and

$$(105.2) \quad A_n(t) = \{x \in X : f_n^*(x) > t\}$$

for each $t > 0$, then

$$(105.3) \quad t \mu(A_n(t)) \leq \int_X \|f_n(x)\| d\mu(x)$$

for every $t > 0$ and $n \geq 1$. This can be shown in the standard way. In particular, one can use the fact that $\{\|f_j\|\}_{j=1}^\infty$ is a submartingale, because of (103.2).

Suppose now that $\|f_n(x)\|$ has uniformly bounded L^1 norm, and put

$$(105.4) \quad f^*(x) = \sup_{j \geq 1} \|f_j(x)\|.$$

If

$$(105.5) \quad A(t) = \{x \in X : f^*(x) > t\}$$

for each $t > 0$, then

$$(105.6) \quad A(t) = \bigcup_{n=1}^\infty A_n(t),$$

and of course $A_n(t) \subseteq A_{n+1}(t)$. It follows that

$$(105.7) \quad t \mu(A(t)) \leq \sup_{n \geq 1} \int_X \|f_n(x)\| d\mu(x)$$

for each $t > 0$, by taking the limit as $n \rightarrow \infty$ in (105.3).

As in Section 85, we can also consider $\{f_j - f_l\}_{j=l}^\infty$ as a V -valued martingale on X with respect to the \mathcal{B}_j 's with $j \geq l$. If

$$(105.8) \quad B_l(t) = \left\{ x \in X : \sup_{j \geq l} \|f_j(x) - f_l(x)\| > t \right\},$$

then we get that

$$(105.9) \quad t \mu(B_l(t)) \leq \sup_{j \geq l} \int_X \|f_j(x) - f_l(x)\| d\mu(x)$$

for each $t > 0$ and $l \geq 1$. Hence

$$(105.10) \quad t \mu\left(\bigcap_{l=1}^\infty B_l(t)\right) \leq \lim_{l \rightarrow \infty} \sup_{j \geq l} \int_X \|f_j(x) - f_l(x)\| d\mu(x)$$

for each $t > 0$.

Suppose that

$$(105.11) \quad \lim_{l \rightarrow \infty} \sup_{j \geq l} \int_X \|f_j(x) - f_l(x)\| d\mu(x) = 0,$$

which means that $\{f_j\}_{j=1}^\infty$ is a Cauchy sequence with respect to the L^1 norm for V -valued functions on X . This together with (105.10) implies that

$$(105.12) \quad \mu\left(\bigcap_{l=1}^\infty B_l(t)\right) = 0$$

for every $t > 0$. Of course,

$$(105.13) \quad X \setminus \left(\bigcap_{l=1}^\infty B_l(t)\right) = \{x \in X : \sup_{j \geq l} \|f_j(x) - f_l(x)\| \leq t \text{ for some } l \in \mathbf{Z}_+\},$$

and it follows that

$$(105.14) \quad \lim_{l \rightarrow \infty} \sup_{j \geq l} \|f_j(x) - f_l(x)\| = 0$$

for almost every $x \in X$, by taking $t = 1/n$ for $n \in \mathbf{Z}_+$. This shows that $\{f_j(x)\}_{j=1}^\infty$ is a Cauchy sequence in V for almost every $x \in X$, and hence that $\{f_j(x)\}_{j=1}^\infty$ converges for almost every $x \in X$, because V is complete. Thus this criterion for convergence almost everywhere works as well in the vector-valued case as for real or complex-valued functions.

106 Another scenario

Let $(X_1, \mathcal{A}_1, \mu_1), (X_2, \mathcal{A}_2, \mu_2), \dots$ be a sequence of probability spaces, and let $X = \prod_{j=1}^\infty X_j$ be their Cartesian product, with the product measure μ . As

usual, let \mathcal{B}_n be the σ -subalgebra of the σ -algebra of measurable subsets of X of the form

$$(106.1) \quad A \times \prod_{j=n+1}^{\infty} X_j,$$

where A is a measurable subset of $\prod_{j=1}^n X_j$. If each X_j has only finitely or countably many elements, and every subset of X_j is measurable, then \mathcal{B}_n consists of the sets of the form (106.1), where A is any subset of $\prod_{j=1}^n X_j$. In this case, \mathcal{B}_n is the σ -algebra generated by the partition \mathcal{P}_n of subsets of X of the form (106.1), where $A \subseteq \prod_{j=1}^n X_j$ has exactly one element.

Let $a_1(x_1), a_2(x_2), \dots$ be a sequence of integrable real or complex-valued functions on X_1, X_2, \dots such that

$$(106.2) \quad \int_{X_j} a_j(x_j) d\mu(x_j) = 0$$

for each j . Also let V be a real or complex vector space with a norm $\|v\|$, and let v_1, v_2, \dots be a sequence of elements of V . Under these conditions, it is natural to consider

$$(106.3) \quad f_n(x) = \sum_{j=1}^n a_j(x_j) v_j$$

as a V -valued martingale on X with respect to the \mathcal{B}_n 's. In this case, it is very easy to understand the meaning of the vector-valued integrals, because of the special form of the functions. This is also consistent with the discussion in Section 103 when the X_j 's have only finitely or countably many elements, and all of their subsets are measurable.

By construction, $f_n(x)$ takes values in a linear subspace of V with dimension less than or equal to n for each $n \in \mathbf{Z}_+$. Thus one can identify f_n with a function on X with values in \mathbf{R}^n or \mathbf{C}^n whose components are measurable. One can also check that $\|f_n(x)\|$ is measurable as a nonnegative real-valued function on X , using the fact that any norm on \mathbf{R}^n or \mathbf{C}^n is bounded by a constant multiple of the standard norm, and hence is continuous with respect to the standard topology. Moreover, $\{\|f_n(x)\|\}_{n=1}^{\infty}$ is a submartingale with respect to the \mathcal{B}_n 's, basically because the norm of the integral of a V -valued function is less than or equal to the integral of the norm of the function.

Suppose that $\|f_n(x)\|$ has uniformly bounded L^1 norm, and let

$$(106.4) \quad f^*(x) = \sup_{n \geq 1} \|f_n(x)\|$$

be the corresponding maximal function. As in the previous section,

$$(106.5) \quad t \mu(\{x \in X : f^*(x) > t\}) \leq \sup_{n \geq 1} \int_X \|f_n\| d\mu$$

for every $t > 0$. This permits one to show that

$$(106.6) \quad \lim_{n \rightarrow \infty} \sup_{l \geq n} \|f_l(x) - f_n(x)\| = 0$$

for almost every $x \in X$ when

$$(106.7) \quad \lim_{n \rightarrow \infty} \sup_{l \geq n} \int_X \|f_l - f_n\| d\mu = 0,$$

as before. Hence $\{f_n(x)\}_{n=1}^\infty$ is a Cauchy sequence in V for almost every x in X under these conditions. If V is complete, then it follows that $\{f_n(x)\}_{n=1}^\infty$ converges for almost every $x \in X$.

107 Hilbert space martingales

Let (X, \mathcal{A}, μ) be a probability space, and suppose that $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \cdots$ is an increasing sequence of σ -subalgebra of \mathcal{A} as in Section 103 or the preceding section. Also let $(V, \langle v, w \rangle)$ be a real or complex Hilbert space, and let $\{f_j\}_{j=1}^\infty$ be a V -valued martingale with respect to the \mathcal{B}_j 's such that $\|f_j(x)\| \in L^2(X)$ for each j .

As in Section 81, one can check that

$$(107.1) \quad \int_X \langle f_j(x), f_{l+1}(x) - f_l(x) \rangle d\mu(x) = 0$$

for each $j \leq l$. If $j < l$, then we get that

$$(107.2) \quad \int_X \langle f_{j+1}(x) - f_j(x), f_{l+1}(x) - f_l(x) \rangle d\mu(x) = 0.$$

Using the identity $f_n = f_1 + \sum_{j=1}^{n-1} (f_{j+1} - f_j)$, it follows that

$$(107.3) \quad \int_X \|f_n(x)\|^2 d\mu(x) = \int_X \|f_1(x)\|^2 d\mu(x) + \sum_{j=1}^{n-1} \int_X \|f_{j+1}(x) - f_j(x)\|^2 d\mu(x)$$

for each n . Similarly,

$$(107.4) \quad \int_X \|f_n(x) - f_l(x)\|^2 d\mu(x) = \sum_{j=l}^{n-1} \int_X \|f_{j+1}(x) - f_j(x)\|^2 d\mu(x)$$

when $n > l$.

If $\|f_n(x)\|$ has bounded L^2 norm, then (107.3) implies that

$$(107.5) \quad \sum_{j=1}^{\infty} \int_X \|f_{j+1}(x) - f_j(x)\|^2 d\mu(x) < \infty.$$

Under these conditions,

$$(107.6) \quad \lim_{l \rightarrow \infty} \sum_{j=l}^{\infty} \int_X \|f_{j+1}(x) - f_j(x)\|^2 d\mu(x) = 0,$$

and hence

$$(107.7) \quad \lim_{l \rightarrow \infty} \sup_{n > l} \int_X \|f_n(x) - f_l(x)\|^2 d\mu(x) = 0.$$

In particular, $\{f_j(x)\}_{j=1}^\infty$ converges in V for almost every $x \in X$, as in the previous sections.

108 Nonnegative submartingales

Let (X, \mathcal{A}, μ) be a probability space, and let $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \dots$ be an increasing sequence of σ -subalgebras of \mathcal{A} . Also let $\{f_j\}_{j=1}^\infty$ be a submartingale with respect to this filtration such that $f_j \geq 0$ for each j . This includes the case of the norm of a vector-valued martingale, as before. If

$$(108.1) \quad f_n^*(x) = \max_{1 \leq j \leq n} f_j(x)$$

and

$$(108.2) \quad A_n(t) = \{x \in X : f_n^*(x) > t\},$$

then

$$(108.3) \quad t \mu(A_n(t)) \leq \int_{A_n(t)} f_n d\mu \leq \int_X f_n d\mu$$

for each $t > 0$ and $n \geq 1$, as shown previously. If

$$(108.4) \quad f^*(x) = \sup_{j \geq 1} f_j(x)$$

and

$$(108.5) \quad A(t) = \{x \in X : f^*(x) > t\},$$

then

$$(108.6) \quad A(t) = \bigcup_{n=1}^{\infty} A_n(t)$$

and

$$(108.7) \quad t \mu(A(t)) \leq \sup_{n \geq 1} \int_X f_n d\mu$$

for each $t > 0$ when the L^1 norms of the f_n 's are bounded.

By hypothesis,

$$(108.8) \quad 0 \leq f_j \leq E(f_n | \mathcal{B}_j)$$

when $j \leq n$, and of course $E(f_n | \mathcal{B}_j)$ is a martingale in j for each n . If $f_n \in L^p(X)$, $1 < p < \infty$, then

$$(108.9) \quad \int_X \left(\max_{1 \leq j \leq n} E(f_n | \mathcal{B}_j) \right)^p d\mu \leq \frac{p 2^{p-1}}{p-1} \int_X f_n^p d\mu,$$

as in Section 88. Hence

$$(108.10) \quad \int_X (f_n^*)^p d\mu \leq \frac{p 2^{p-1}}{p-1} \int_X f_n^p d\mu.$$

If the L^p norm of f_n is uniformly bounded in n , then the monotone convergence theorem implies that $f^* \in L^p$, with

$$(108.11) \quad \int_X (f^*)^p d\mu \leq \frac{p 2^{p-1}}{p-1} \sup_{n \geq 1} \int_X f_n^p d\mu.$$

Thus one gets the same L^p estimates for nonnegative submartingales as for martingales.

109 L^p -Valued martingales

As in Section 104, we can look at L^p -valued martingales in terms of functions on a product space. Let (X, \mathcal{A}, μ) be a probability space, and let $\mathcal{P}_1, \mathcal{P}_2, \dots$ be a sequence of partitions of X into finitely or countably many measurable subsets with positive measure such that \mathcal{P}_{j+1} is a refinement of \mathcal{P}_j for each j . Also let (Y, \mathcal{B}, ν) be a σ -finite measure space, and fix p , $1 < p < \infty$.

If $f_l(x)$ is an $L^p(Y)$ -valued function on X that is constant on the elements of \mathcal{P}_l , then

$$(109.1) \quad F_l(x, y) = f_l(x)(y)$$

is a function on $X \times Y$ that is constant in x on each element of \mathcal{P}_l and measurable in y for each $x \in X$. If $\|f_l(x)\|_{L^p(Y)} \in L^p(X)$, then $F_l(x, y) \in L^p(X \times Y)$, and

$$(109.2) \quad \begin{aligned} \int_X \|f_l(x)\|_{L^p(Y)}^p d\mu(x) &= \int_X \left(\int_Y |F_l(x, y)|^p d\nu(y) \right) d\mu(x) \\ &= \int_{X \times Y} |F_l(x, y)|^p d(\mu \times \nu)(x, y). \end{aligned}$$

Conversely, if $F_l(x, y) \in L^p(X \times Y)$ is constant in x on each element of \mathcal{P}_l , then we get an $L^p(Y)$ -valued function $f_l(x)$ on X that is constant on each element of \mathcal{P}_l and for which $\|f_l(x)\|_{L^p(Y)} \in L^p(X)$. If $\widehat{\mathcal{B}}_l$ is the σ -algebra of subsets of $X \times Y$ that corresponds to $\mathcal{B}_l = \mathcal{B}(\mathcal{P}_l)$ on X and \mathcal{B} on Y as before, then $F_l(x, y)$ is measurable with respect to $\widehat{\mathcal{B}}_l$.

Now let $\{f_l\}_{l=1}^\infty$ be a sequence of $L^p(Y)$ -valued functions on X such that $f_l(x)$ is constant on each element of \mathcal{P}_l and $\|f_l(x)\|_{L^p(Y)} \in L^p(X)$ for each l . This corresponds exactly to a sequence of functions $\{F_l\}_{l=1}^\infty$ in $L^p(X \times Y)$ such that $F_l(x, y)$ is measurable with respect to $\widehat{\mathcal{B}}_l$ for each l , as in the preceding paragraph. If $\{f_l\}_{l=1}^\infty$ is an $L^p(Y)$ -valued martingale on X with respect to the \mathcal{B}_l 's and Y is a probability space, then $X \times Y$ is also a probability space, and $\{F_l\}_{l=1}^\infty$ is a martingale on $X \times Y$ with respect to the $\widehat{\mathcal{B}}_l$'s. If the $L^p(X)$ norm of $\|f_l(x)\|_{L^p(Y)}$ is bounded, then the $L^p(X \times Y)$ norm of $F_l(x, y)$ is bounded, and hence $\{F_l\}_{l=1}^\infty$ converges in $L^p(X \times Y)$. In particular, $\{F_l\}_{l=1}^\infty$ is a Cauchy sequence in $L^p(X \times Y)$, which implies that

$$(109.3) \quad \lim_{l \rightarrow \infty} \sup_{j \geq l} \int_X \|f_j(x) - f_l(x)\|_{L^p(Y)}^p d\mu(x) = 0.$$

Of course, the same conclusion holds when $0 < \nu(Y) < \infty$, by dividing by $\nu(Y)$ to get a probability space. Otherwise, let ρ be a strictly positive measurable function on Y such that

$$(109.4) \quad \int_Y \rho(y) d\nu(y) = 1,$$

which is possible because (Y, \mathcal{B}, ν) is supposed to be σ -finite. Thus

$$(109.5) \quad \nu_\rho(B) = \int_B \rho(y) d\nu(y)$$

is a probability measure on (Y, \mathcal{B}) . If $\phi(y) \in L^p(Y, \nu)$, then

$$(109.6) \quad \phi_\rho(y) = \phi(y) \rho(y)^{-1/p} \in L^p(Y, \nu_\rho),$$

and

$$(109.7) \quad \int_Y |\phi_\rho(y)|^p d\nu_\rho(y) = \int_Y |\phi(y)|^p d\nu(y).$$

Using this, one can check that (109.3) holds for any σ -finite measure space (Y, \mathcal{B}, ν) , by reducing to the probability space $(Y, \mathcal{B}, \nu_\rho)$.

110 Another criterion

Let (X, \mathcal{A}, μ) be a probability space, and let $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \dots$ be an increasing sequence of σ -subalgebras of \mathcal{A} as in Section 103 or 106. Also let V be a real or complex Banach space with a norm $\|v\|$, and let $\{f_j\}_{j=1}^\infty$ be a V -valued martingale on X with respect to the \mathcal{B}_j 's such that $\|f_j(x)\| \in L^1(X)$ for each j . Suppose that for each $\epsilon > 0$ there is a V -valued martingale $\{g_j\}_{j=1}^\infty$ on X with respect to the \mathcal{B}_j 's such that

$$(110.1) \quad \int_X \|f_j(x) - g_j(x)\| d\mu(x) \leq \epsilon$$

for each j , and $\{g_j(x)\}_{j=1}^\infty$ converges in V for almost every $x \in X$. Let us check that $\{f_j(x)\}_{j=1}^\infty$ converges in V for almost every $x \in X$ under these conditions.

Of course, it suffices to show that

$$(110.2) \quad \lim_{l \rightarrow \infty} \sup_{j \geq l} \|f_j(x) - f_l(x)\| = 0$$

for almost every $x \in X$, so that $\{f_j(x)\}_{j=1}^\infty$ is a Cauchy sequence in V for almost every $x \in X$. Put $h_j = f_j - g_j$, so that $\{h_j\}_{j=1}^\infty$ is also a V -valued martingale on X with respect to the \mathcal{B}_j 's. Observe that

$$(110.3) \quad \begin{aligned} \lim_{l \rightarrow \infty} \sup_{j \geq l} \|f_j(x) - f_l(x)\| &\leq \lim_{l \rightarrow \infty} \sup_{j \geq l} \|g_j(x) - g_l(x)\| \\ &\quad + \lim_{l \rightarrow \infty} \sup_{j \geq l} \|h_j(x) - h_l(x)\| \end{aligned}$$

for every $x \in X$. This implies that

$$(110.4) \quad \lim_{l \rightarrow \infty} \sup_{j \geq l} \|f_j(x) - f_l(x)\| \leq \lim_{l \rightarrow \infty} \sup_{j \geq l} \|h_j(x) - h_l(x)\|$$

for almost every $x \in X$, because $\{g_j(x)\}_{j=1}^\infty$ is a Cauchy sequence in V for almost every $x \in X$. Hence

$$(110.5) \quad \lim_{l \rightarrow \infty} \sup_{j \geq l} \|f_j(x) - f_l(x)\| \leq 2 \sup_{j \geq 1} \|h_j(x)\| = 2h^*(x)$$

for almost every $x \in X$.

By the usual maximal function estimate,

$$(110.6) \quad t \mu(\{x \in X : h^*(x) > t\}) \leq \sup_{j \geq 1} \int_X \|h_j(x)\| d\mu(x) \leq \epsilon$$

for every $t > 0$. If

$$(110.7) \quad E(t) = \left\{ x \in X : \lim_{l \rightarrow \infty} \sup_{j \geq l} \|f_j(x) - f_l(x)\| > 2t \right\},$$

then

$$(110.8) \quad \mu(E(t)) \leq \mu(\{x \in X : h^*(x) > t\}),$$

by (110.5), and so

$$(110.9) \quad t \mu(E(t)) \leq \epsilon$$

for every $\epsilon, t > 0$. Because $E(t)$ does not depend on ϵ , we may conclude that $\mu(E(t)) = 0$ for every $t > 0$. This implies that (110.2) holds for almost every $x \in X$, as desired.

Note that this criterion is satisfied when

$$(110.10) \quad \lim_{l \rightarrow \infty} \sup_{j \geq l} \int_X \|f_j(x) - f_l(x)\| d\mu(x) = 0.$$

To see this, one can take $g_j(x)$ to be of the form $f_{\min(j, N)}(x)$ for large positive integers N . This converges as $j \rightarrow \infty$ for each fixed N trivially, and (110.1) holds for sufficiently large N by hypothesis.

111 ℓ^1 -Valued martingales

As before, let (X, \mathcal{A}, μ) be a probability space, and let $\mathcal{P}_1, \mathcal{P}_2, \dots$ be a sequence of partitions of X into finitely or countably many measurable subsets with positive measure such that \mathcal{P}_{j+1} is a refinement of \mathcal{P}_j for each j . Suppose that $\{f_j\}_{j=1}^\infty$ is a sequence of functions on X with values in $\ell^1 = \ell^1(\mathbf{Z}_+)$. Thus for each $x \in X$ and $j \geq 1$ we get a summable sequence $\{f_{j,k}(x)\}_{k=1}^\infty$ of real or complex numbers. Of course, $f_j(x)$ is constant on each element of \mathcal{P}_j if and only if $f_{j,k}(x)$

is constant on each element of \mathcal{P}_j for each $k \geq 1$. If $\|f_j(x)\|_{\ell^1}$ is integrable on X , then $f_{j,k}(x)$ is integrable on X for each k , and

$$(111.1) \quad \int_X \|f_j(x)\|_{\ell^1} d\mu(x) = \int_X \sum_{k=1}^{\infty} |f_{j,k}(x)| d\mu(x) = \sum_{k=1}^{\infty} \int_X |f_{j,k}(x)| d\mu(x).$$

Suppose now that $\{f_j\}_{j=1}^{\infty}$ is an ℓ^1 -valued martingale on X with respect to $\mathcal{B}_j = \mathcal{B}(\mathcal{P}_j)$. This implies that $\{f_{j,k}\}_{j=1}^{\infty}$ is a martingale on X with respect to the \mathcal{B}_j 's for each k . In particular,

$$(111.2) \quad \int_X |f_{j,k}(x)| d\mu(x) \leq \int_X |f_{j+1,k}(x)| d\mu(x)$$

for each $j, k \geq 1$, and hence

$$(111.3) \quad \int_X \|f_j(x)\|_{\ell^1} d\mu(x) \leq \int_X \|f_{j+1}(x)\|_{\ell^1} d\mu(x)$$

for each j .

Suppose also that the $L^1(X)$ norm of $\|f_j(x)\|_{\ell^1}$ is bounded. Because of monotonicity,

$$(111.4) \quad \sup_{j \geq 1} \int_X \|f_j(x)\|_{\ell^1} d\mu(x) = \lim_{j \rightarrow \infty} \int_X \|f_j(x)\|_{\ell^1} d\mu(x),$$

and similarly

$$(111.5) \quad \sup_{j \geq 1} \int_X |f_{j,k}(x)| d\mu(x) = \lim_{j \rightarrow \infty} \int_X |f_{j,k}(x)| d\mu(x)$$

for each k . The monotone convergence theorem for sums implies that

$$(111.6) \quad \lim_{j \rightarrow \infty} \int_X \|f_j(x)\|_{\ell^1} d\mu(x) = \sum_{k=1}^{\infty} \left(\lim_{j \rightarrow \infty} \int_X |f_{j,k}(x)| d\mu(x) \right).$$

Therefore

$$(111.7) \quad \sup_{j \geq 1} \int_X \|f_j(x)\|_{\ell^1} d\mu(x) = \sum_{k=1}^{\infty} \left(\sup_{j \geq 1} \int_X |f_{j,k}(x)| d\mu(x) \right).$$

Let N be a large positive integer, and put

$$(111.8) \quad g_{j,k}(x) = f_{j,k}(x), \quad h_{j,k}(x) = 0 \quad \text{when } k \leq N,$$

$$(111.9) \quad g_{j,k}(x) = 0, \quad h_{j,k}(x) = f_{j,k}(x) \quad \text{when } k > N.$$

If $g_j(x) = \{g_{j,k}(x)\}_{k=1}^{\infty}$, $h_j(x) = \{h_{j,k}(x)\}_{k=1}^{\infty}$, then $g_j(x), h_j(x) \in \ell^1$ and

$$(111.10) \quad f_j(x) = g_j(x) + h_j(x).$$

Note that $\{g_j(x)\}_{j=1}^{\infty}$ converges for almost every $x \in X$, as a consequence of the convergence almost everywhere of real or complex martingales with bounded L^1 norm. One can also check that $\|h_j(x)\|_{\ell^1}$ has small $L^1(X)$ norm, uniformly in j , and for sufficiently large N , by the discussion in the preceding paragraph. Thus $\{f_j\}_{j=1}^{\infty}$ satisfies the criterion described in the previous section, and it follows that $\{f_j(x)\}_{j=1}^{\infty}$ converges in ℓ^1 for almost every $x \in X$.

112 Differentiability of paths

Let $(V, \|v\|)$ be a real or complex Banach space, and let $f : [a, b] \rightarrow V$ be a path of finite length. Suppose that for each $\epsilon > 0$ there is a path $g : [a, b] \rightarrow V$ of finite length such that the length of $f - g$ on $[a, b]$ is less than or equal to ϵ and g is differentiable almost everywhere on $[a, b]$. We would like to show that f is also differentiable almost everywhere on $[a, b]$.

If $a \leq x \leq b$ and $r > 0$, then let $\delta_r(f)(x)$ be the set of difference quotients

$$(112.1) \quad \frac{f(x) - f(y)}{x - y},$$

where $a \leq y \leq b$ and $0 < |x - y| < r$. One can check that f is differentiable at x if and only if

$$(112.2) \quad \lim_{r \rightarrow 0} \text{diam } \delta_r(f)(x) = 0,$$

using the completeness of V for the “if” part. Put $h = f - g$, and observe that

$$(112.3) \quad \text{diam } \delta_r(f)(x) \leq \text{diam } \delta_r(g)(x) + \text{diam } \delta_r(h)(x)$$

for every $x \in [a, b]$ and $r > 0$.

By hypothesis,

$$(112.4) \quad \lim_{r \rightarrow 0} \text{diam } \delta_r(g)(x) = 0$$

for almost every $x \in [a, b]$. Hence

$$(112.5) \quad \lim_{r \rightarrow 0} \delta_r(f)(x) \leq \sup_{r > 0} \text{diam } \delta_r(h)(x)$$

for almost every $x \in [a, b]$.

Using maximal functions as in Section 50, we get that for each $t > 0$ there is an open set $E_t(h) \subseteq \mathbf{R}$ such that

$$(112.6) \quad \|h(x) - h(y)\| \leq t \|x - y\|$$

when $E_t(h)$ does not contain the interval connecting $x, y \in [a, b]$, and

$$(112.7) \quad |E_t(h)| \leq 2 \epsilon t^{-1}.$$

Here $|E_t(h)|$ denotes the Lebesgue measure of $E_t(h)$, as usual. Thus

$$(112.8) \quad \sup_{r > 0} \delta_r(h)(x) \leq 2 t$$

for every $x \in [a, b] \setminus E_t(h)$. It follows that

$$(112.9) \quad \lim_{r \rightarrow 0} \delta_r(f)(x) \leq 2 t$$

for almost every $x \in [a, b] \setminus E_t(h)$. Using these estimates for every $\epsilon, t > 0$, we get that (112.2) holds for almost every $x \in [a, b]$, as desired.

113 Paths in ℓ^1

Let $f : [a, b] \rightarrow \ell^1 = \ell^1(\mathbf{Z}_+)$ be a path of finite length. Thus $f(x) = \{f_j(x)\}_{j=1}^\infty$, where each f_j is a real or complex-valued function on $[a, b]$ of bounded variation. More precisely, let l be a positive integer, and let $\mathcal{P}_1, \dots, \mathcal{P}_l$ be partitions of $[a, b]$, as in Section 41. Also let \mathcal{P} be a partition of $[a, b]$ that is a common refinement of $\mathcal{P}_1, \dots, \mathcal{P}_l$. If $\Lambda_a^b(f, \mathcal{P})$ denotes the approximation to the length of f associated to \mathcal{P} , and similarly for the f_j 's and \mathcal{P}_j 's, then

$$(113.1) \quad \sum_{j=1}^l \Lambda_a^b(f_j, \mathcal{P}_j) \leq \sum_{j=1}^l \Lambda_a^b(f_j, \mathcal{P}) \leq \Lambda_a^b(f, \mathcal{P}).$$

Hence

$$(113.2) \quad \sum_{j=1}^l \Lambda_a^b(f_j, \mathcal{P}_j) \leq \Lambda_a^b(f),$$

where $\Lambda_a^b(f)$ denotes the length of f on $[a, b]$. This implies that

$$(113.3) \quad \sum_{j=1}^l \Lambda_a^b(f_j) \leq \Lambda(f).$$

because $\mathcal{P}_1, \dots, \mathcal{P}_l$ are arbitrary partitions of $[a, b]$. Therefore

$$(113.4) \quad \sum_{j=1}^\infty \Lambda_a^b(f_j) \leq \Lambda(f),$$

because $l \geq 1$ is arbitrary.

Similarly, if \mathcal{P} is any partition of $[a, b]$, then

$$(113.5) \quad \Lambda_a^b(f, \mathcal{P}) = \sum_{j=1}^\infty \Lambda_a^b(f_j, \mathcal{P}_j) \leq \sum_{j=1}^\infty \Lambda_a^b(f_j).$$

This implies that

$$(113.6) \quad \Lambda_a^b(f) \leq \sum_{j=1}^\infty \Lambda_a^b(f_j).$$

It follows that

$$(113.7) \quad \Lambda_a^b(f) = \sum_{j=1}^\infty \Lambda_a^b(f_j),$$

by the remarks in the preceding paragraph.

Let N be a large positive integer, and put

$$(113.8) \quad \begin{aligned} g_j(x) &= f_j(x), \quad h_j(x) = 0 & \text{when } j \leq N, \\ g_j(x) &= 0, \quad h_j(x) = f_j(x) & \text{when } j > N. \end{aligned}$$

If $g(x) = \{g_j(x)\}_{j=1}^\infty$, $h(x) = \{h_j(x)\}_{j=1}^\infty$, then $g(x), h(x) \in \ell^1$ for every x in $[a, b]$, and

$$(113.9) \quad f(x) = g(x) + h(x).$$

Observe that $f, g : [a, b] \rightarrow \ell^1$ have finite length, and that the length of h on $[a, b]$ tends to 0 as $N \rightarrow \infty$. We also know that g is differentiable almost everywhere on $[a, b]$, by the corresponding results for real or complex-valued functions. It follows that f is also differentiable almost everywhere on $[a, b]$, as in the previous section.

114 L^p -Valued functions

Let (Y, \mathcal{B}, ν) be a σ -finite measure space, and consider $\mathbf{R} \times Y$, equipped with the product measure corresponding to Lebesgue measure on the real line. A function $F(x, y) \in L^p(\mathbf{R} \times Y)$, $1 \leq p < \infty$, may be considered as representing an L^p function on \mathbf{R} with values in $L^p(Y)$. Put

$$(114.1) \quad F_p(x) = \left(\int_Y |F(x, y)|^p d\nu(y) \right)^{1/p},$$

which is the $L^p(Y)$ norm of $F(x, y)$ in y . By Fubini's theorem,

$$(114.2) \quad \left(\int_{\mathbf{R} \times Y} |F(x, y)|^p dx d\nu(y) \right)^{1/p} = \left(\int_{\mathbf{R}} F_p(x)^p dx \right)^{1/p}.$$

Thus the $L^p(\mathbf{R} \times Y)$ norm of $F(x, y)$ is the same as starting with the $L^p(Y)$ norm of $F(x, y)$ in y , and then taking the $L^p(\mathbf{R})$ norm of the result in x .

Put

$$(114.3) \quad L(F)(x) = \limsup_{r \rightarrow 0} \frac{1}{2r} \int_{x-r}^{x+r} \left(\int_Y |F(t, y) - F(x, y)|^p d\nu(y) \right)^{1/p} dt.$$

As in Section 47, we would like to say that

$$(114.4) \quad L(F)(x) = 0$$

for almost every $x \in \mathbf{R}$. As usual, there are two main ingredients in the proof. The first is that this condition holds for a dense class of functions $F(x, y)$ in $L^p(x, y)$. In this case, one can use finite linear combinations of functions of the form $f(x)g(y)$, where $f(x) \in L^p(\mathbf{R})$ and $g(y) \in L^p(Y)$. If one also takes $f(x)$ to be continuous, then the limit is equal to 0 at every $x \in \mathbf{R}$. If Y is a locally compact Hausdorff topological space and ν is a Borel measure on Y with suitable regularity properties, then one can use continuous functions on $\mathbf{R} \times Y$ with compact support as the dense class. Again the limit is equal to 0 for every $x \in \mathbf{R}$ in this situation. The second main ingredient is an estimate for an appropriate maximal function, which reduces here to the Hardy-Littlewood maximal function of $F_p(x) \in L^p(\mathbf{R})$.

115 Continuous L^p -valued functions

Let (Y, \mathcal{B}, ν) be a measure space, and let f be a continuous function on the real line with values in $L^p(Y)$, $1 \leq p < \infty$. If $g \in L^p(Y)$, then

$$(115.1) \quad \{y \in Y : |g(y)| \geq 1/n\}$$

has finite measure for each $n \in \mathbf{Z}_+$, and hence

$$(115.2) \quad \{y \in Y : g(y) \neq 0\}$$

is σ -finite. Applying this to $f(r)$ for each rational number r , we get that there is a σ -finite measurable set $Y_0 \subseteq Y$ such that $f(r) = 0$ on $Y \setminus Y_0$ for every $r \in \mathbf{Q}$. This implies that $f(r) = 0$ almost everywhere on $Y \setminus Y_0$ for every $r \in \mathbf{R}$, because $f : \mathbf{R} \rightarrow L^p(Y)$ is continuous. Thus we may as well suppose that Y is σ -finite.

Let us now restrict our attention to the case where f has compact support on \mathbf{R} . More precisely, let $I = [a, b]$ be a closed interval in the real line such that $f(x) = 0$ when $x \in \mathbf{R} \setminus [a, b]$. Consider the product $\mathbf{R} \times L^p(Y)$ with the product measure associated to Lebesgue measure on \mathbf{R} , as in the preceding section. We would like to check that there is an $F(x, y) \in L^p(\mathbf{R} \times Y)$ such that

$$(115.3) \quad f(x) = F(x, \cdot)$$

as elements of $L^p(Y)$ for almost every $x \in \mathbf{R}$. In this case,

$$(115.4) \quad \begin{aligned} \int_{\mathbf{R} \times Y} |F(x, y)|^p dx d\nu(y) &= \int_{\mathbf{R}} \left(\int_Y |F(x, y)|^p d\nu(y) \right) dx \\ &= \int_{\mathbf{R}} \|f(x)\|_{L^p(Y)}^p dx. \end{aligned}$$

In particular, the $L^p(\mathbf{R} \times Y)$ norm of $F(x, y)$ would be bounded by a constant multiple of the supremum norm of $\|f(x)\|_{L^p(Y)}$ on I . If $F(x, y), \tilde{F}(x, y) \in L^p(\mathbf{R} \times Y)$ both satisfy (115.3) for almost every $x \in \mathbf{R}$, then it follows that $F(x, y) = \tilde{F}(x, y)$ for almost every $(x, y) \in \mathbf{R} \times Y$.

If $f(x) = \phi(x)g$ for some real or complex-valued function $\phi(x)$ with compact support on \mathbf{R} and some $g \in L^p(Y)$, then we can simply take $F(x, y) = \phi(x)g(y)$. Similarly, if $f(x)$ is a finite linear combination of $L^p(Y)$ -valued functions on \mathbf{R} of this form, then it is easy to get $F(x, y)$. Otherwise, one can approximate $f(x)$ by a sequence $\{f_j(x)\}_{j=1}^\infty$ of $L^p(Y)$ -valued functions of this type with respect to the supremum norm of $\|f(x)\|_{L^p(Y)}$ on I . By construction, $f_j(x)$ corresponds to a function $F_j(x, y)$ in $L^p(\mathbf{R} \times Y)$ for each j . Moreover, $\{F_j(x, y)\}_{j=1}^\infty$ is a Cauchy sequence in $L^p(\mathbf{R} \times Y)$, because of (115.4). Hence $\{F_j(x, y)\}_{j=1}^\infty$ converges to a function $F(x, y)$ in $L^p(\mathbf{R} \times Y)$. It is not too difficult to verify that this function $F(x, y)$ satisfies (115.3), as desired.

116 Lipschitz L^p -valued functions

Let (Y, \mathcal{B}, ν) be a σ -finite measure space, and suppose that $f : \mathbf{R} \rightarrow L^p(Y)$ is a Lipschitz mapping for some $1 < p < \infty$. It will be convenient to ask also

at first that $f(x)$ have compact support in \mathbf{R} , which is to say that there is a closed interval $[a, b]$ in the real line such that $f(x) = 0$ when $x \in \mathbf{R} \setminus [a, b]$. Let $F(x, y)$ be the function in $L^p(\mathbf{R} \times Y)$ that corresponds to $f(x)$ as in the previous section. Because $f(x)$ has compact support, the ordinary Lipschitz condition implies an integrated Lipschitz condition of the form

$$(116.1) \quad \left(\int_{\mathbf{R}} \|f(x+h) - f(x)\|_{L^p(Y)}^p dx \right)^{1/p} \leq C|h|$$

for some $C \geq 0$ and every $h \in \mathbf{R}$. This implies that

$$(116.2) \quad \left(\int_{\mathbf{R} \times Y} |F(x+h, y) - F(x, y)|^p dx d\nu(y) \right)^{1/p} \leq C|h|.$$

Let q be the exponent conjugate to p , so that $1/p + 1/q = 1$. If $h \in \mathbf{R}$, $h \neq 0$, and $\Phi(x, y) \in L^q(\mathbf{R} \times Y)$, then put

$$(116.3) \quad \lambda_h(\Phi) = \int_{\mathbf{R} \times Y} \frac{F(x+h, y) - F(x, y)}{h} \Phi(x, y) dx d\nu(y).$$

This defines a bounded linear functional on $L^q(\mathbf{R} \times Y)$, with dual norm less than or equal to C , by Hölder's inequality. We also have that

$$(116.4) \quad \lambda_h(\Phi) = - \int_{\mathbf{R} \times Y} F(x, y) \frac{\Phi(x, y) - \Phi(x-h, y)}{h} dx d\nu(y),$$

using the change of variables $x \mapsto x - h$. If

$$(116.5) \quad \Phi(x, y) = \phi(x) \psi(y),$$

where $\phi(x)$ is a continuously-differentiable real or complex-valued function on the real line with compact support and $\psi(y) \in L^q(Y)$, then we get that

$$(116.6) \quad \lim_{h \rightarrow 0} \lambda_h(\Phi) = - \int_{\mathbf{R} \times Y} F(x, y) \phi'(x) \psi(y) dx d\nu(y).$$

Similarly,

$$(116.7) \quad \lim_{h \rightarrow 0} \lambda_h(\Phi)$$

exists when $\Phi(x, y)$ is a finite linear combination of functions of this form. As in Section 52, it follows that (116.7) exists for all $\Phi(x, y) \in L^q(\mathbf{R} \times Y)$, since it exists for a dense linear subspace of $L^q(\mathbf{R} \times Y)$, and since the dual norms of λ_h , $h \in \mathbf{R} \setminus \{0\}$, are bounded.

Thus (116.7) defines a bounded linear functional on $L^q(\mathbf{R} \times Y)$. By the Riesz representation theorem, there is a function $G(x, y)$ in $L^p(\mathbf{R} \times Y)$ such that

$$(116.8) \quad \lim_{h \rightarrow 0} \lambda_h(\Phi) = \int_{\mathbf{R} \times Y} G(x, y) \Phi(x, y) dx d\nu(y)$$

for every $\Phi(x, y) \in L^q(\mathbf{R} \times Y)$. In particular,

$$(116.9) \quad \int_{\mathbf{R} \times Y} F(x, y) \phi'(x) \psi(y) dx d\nu(y) = \\ - \int_{\mathbf{R} \times Y} G(x, y) \phi(x) \psi(y) dx d\nu(y)$$

when $\phi(x)$ is a real or complex-valued continuously-differentiable function on the real line with compact support and $\psi(y) \in L^q(Y)$. Put

$$(116.10) \quad f_\psi(x) = \int_Y f(x)(y) \psi(y) d\nu(y)$$

for each $x \in \mathbf{R}$. More precisely, $f(x) \in L^p(Y)$ for every $x \in \mathbf{R}$, and $f_\psi(x)$ is the integral of the product of this function with $\psi \in L^q(Y)$ over Y . Thus $f_\psi(x)$ is a Lipschitz function on \mathbf{R} with compact support for every $\psi \in L^q(Y)$, because $f : \mathbf{R} \rightarrow L^p(Y)$ is a Lipschitz mapping with compact support. Using (116.9), we get that

$$(116.11) \quad \int_{\mathbf{R}} f_\psi(x) \phi'(x) dx = - \int_{\mathbf{R} \times Y} G(x, y) \phi(x) \psi(y) dx d\nu(y)$$

for every $\phi(x), \psi(y)$ as before. This implies that

$$(116.12) \quad f'_\psi(x) = \int_Y G(x, y) \psi(y) d\nu(y)$$

for every $\psi \in L^q(Y)$ in the sense of distributions, as in Section 56. Hence

$$(116.13) \quad f_\psi(t) - f_\psi(r) = \int_r^t \int_Y G(x, y) \psi(y) d\nu(y) dx$$

for every $r, t \in \mathbf{R}$ with $r < t$ and $\psi \in L^q(Y)$. It follows that

$$(116.14) \quad f(t) - f(r) = \int_r^t G(x, \cdot) dx$$

when $r < t$, where both sides of the equation are elements of $L^p(Y)$.

Now that we have this expression for differences of the values of f , one can use the analogue of Lebesgue's theorem in this context to conclude that f is differentiable almost everywhere as an $L^p(Y)$ -valued function on the real line. This works as well for Lipschitz mappings from the real line into $L^p(Y)$ that may not have compact support, since the problem is local. This also works for paths of finite length in $L^p(Y)$, $1 < p < \infty$, because of the approximation arguments in Sections 50 and 51.

117 More duality

Let (Y, \mathcal{B}, ν) be a σ -finite measure space, and let f be a continuous function from the real line into $L^p(Y)$, $1 < p < \infty$. Suppose also that f has compact support in \mathbf{R} , and let q be the exponent conjugate to p , so that $1/p + 1/q = 1$. We would like to define a bounded linear functional on $L^q(\mathbf{R} \times Y)$ directly by

$$(117.1) \quad \Lambda(\Phi) = \int_{\mathbf{R}} \left(\int_Y f(x)(y) \Phi(x, y) d\nu(y) \right) dx.$$

Because of Hölder's inequality,

$$(117.2) \quad \left| \int_Y f(x)(y) \Phi(x, y) d\nu(y) \right| \leq \|f(x)\|_{L^p(Y)} \left(\int_Y |\Phi(x, y)|^q d\nu(y) \right)^{1/q}$$

and

$$(117.3) \quad \begin{aligned} & \int_{\mathbf{R}} \|f(x)\|_{L^p(Y)} \left(\int_Y |\Phi(x, y)|^q d\nu(y) \right)^{1/q} dx \\ & \leq \left(\int_{\mathbf{R}} \|f(x)\|_{L^p(Y)}^p dx \right)^{1/p} \left(\int_{\mathbf{R} \times Y} |\Phi(x, y)|^q dx d\nu(y) \right)^{1/q}. \end{aligned}$$

However, one should be a bit careful about the measurability of

$$(117.4) \quad \int_Y f(x)(y) \Phi(x, y) d\nu(y)$$

as a function of x . If $\Phi(x, y) = \phi(x) \psi(y)$ for some $\phi(x) \in L^q(\mathbf{R})$, $\psi(y) \in L^q(Y)$, then this reduces to

$$(117.5) \quad \phi(x) \int_Y f(x)(y) \psi(y) d\nu(y).$$

The continuity of $f : \mathbf{R} \rightarrow L^p(Y)$ implies that

$$(117.6) \quad \int_Y f(x)(y) \psi(y) d\nu(y)$$

is continuous in x , and so there is no problem in this case. Because linear combinations of functions of this type are dense in $L^q(\mathbf{R} \times Y)$, one can use this to extend $\Lambda(\Phi)$ to all $\Phi \in L^q(\mathbf{R} \times Y)$. Similarly,

$$(117.7) \quad \lambda_h(\Phi) = \int_{\mathbf{R}} \left(\int_Y \frac{f(x+h)(y) - f(x)(y)}{h} \Phi(x, y) d\nu(y) \right) dx$$

can be defined more directly as a bounded linear functional on $L^q(\mathbf{R} \times Y)$ for each $h \in \mathbf{R} \setminus \{0\}$. Equivalently,

$$(117.8) \quad \lambda_h(\Phi) = - \int_{\mathbf{R}} \left(\int_Y f(x)(h) \frac{\Phi(x, y) - \Phi(x-h, y)}{h} d\nu(y) \right) dx.$$

If $\Phi(x, y) = \phi(x) \psi(y)$, where now $\phi(x)$ is a continuously-differentiable function on \mathbf{R} with compact support and $\psi(y) \in L^q(Y)$, then it follows that

$$(117.9) \quad \lim_{h \rightarrow 0} \lambda_h(\Phi) = - \int_{\mathbf{R}} \phi'(x) \left(\int_Y f(x)(y) \psi(y) d\nu(y) \right) dx.$$

At this point, one can continue as in the preceding section when $f : \mathbf{R} \rightarrow L^p(Y)$ is Lipschitz.

118 ℓ^p -Valued functions

Of course, the arguments in the previous sections can be simplified when the functions take values in $\ell^p = \ell^p(\mathbf{Z}_+)$, and there are some commonalities with $p = 1$. Suppose that $f(x) = \{f_j(x)\}_{j=1}^\infty$ is a Lipschitz function on the real line with values in ℓ^p , $1 \leq p \leq \infty$. In particular, $f_j(x)$ is a Lipschitz function on \mathbf{R} for each j , and hence is differentiable almost everywhere. Using the Lipschitz condition for $f : \mathbf{R} \rightarrow \ell^p$, one can check that $\{f'_j(x)\}_{j=1}^\infty \in \ell^p$ for every $x \in \mathbf{R}$ such that $f'_j(x)$ exists for each j , with ℓ^p norm bounded by the Lipschitz constant for f . We also have that

$$(118.1) \quad f_j(t) - f_j(r) = \int_r^t f'_j(x) dx$$

for every $r, t \in \mathbf{R}$ with $r < t$. If $p < \infty$, then one can use this to show that f is differentiable almost everywhere on \mathbf{R} as a mapping into ℓ^p , with derivative given by $\{f'_j(x)\}_{j=1}^\infty$. As usual, it is convenient to restrict one's attention initially to functions f with compact support, so that $\|\{f'_j(x)\}\|_{\ell^p} \in L^p(\mathbf{R})$. As in the $p = 1$ case, one can approximate f by functions with only finitely many nonzero components, for which differentiability almost everywhere is already known. One can then use maximal function estimates to show that the errors are small most of the time.

Note that a Lipschitz mapping from the real line into a separable Hilbert space is differentiable almost everywhere, by the $p = 2$ case. This can be extended to paths of finite length in a separable Hilbert space, because of the approximation arguments in Sections 50 and 51. As in Section 45, any path of finite length is continuous at all but finitely or countably many elements of its domain, and hence is contained in a separable subspace of the range. This implies that a path of finite length in any Hilbert space is differentiable almost everywhere, because it is contained in a separable Hilbert subspace.

119 Products and σ -subalgebras

Let $(X_1, \mathcal{A}_1, \mu_1)$, $(X_2, \mathcal{A}_2, \mu_2)$ be probability spaces, and let $X = X_1 \times X_2$ be their Cartesian product, with the product measure $\mu = \mu_1 \times \mu_2$. Also let \mathcal{B}_1 , \mathcal{B}_2 be σ -subalgebras of \mathcal{A}_1 , \mathcal{A}_2 , respectively, and let \mathcal{B} be the corresponding σ -subalgebra of the σ -algebra of measurable subsets of X . If $\phi_1(x_1) \in L^1(X_1)$,

$\phi_2(x_2) \in L^1(X_2)$, then $\phi(x_1, x_2) = \phi_1(x_1)\phi_2(x_2) \in L^1(X)$, and we would like to check that

$$(119.1) \quad E_X(\phi \mid \mathcal{B}) = E_{X_1}(\phi_1 \mid \mathcal{B}_1) E_{X_2}(\phi_2 \mid \mathcal{B}_2),$$

where the subscripts of E are included to indicate the spaces on which the conditional expectations are taken. Both sides of the equation are measurable with respect to \mathcal{B} , and so it suffices to verify that

$$(119.2) \quad \int_B E_X(\phi \mid \mathcal{B}) d\mu = \int_B E_{X_1}(\phi_1 \mid \mathcal{B}_1) E_{X_2}(\phi_2 \mid \mathcal{B}_2) d\mu$$

for every $B \in \mathcal{B}$. This reduces to

$$(119.3) \quad \int_B \phi d\mu = \int_B E_{X_1}(\phi_1 \mid \mathcal{B}_1) E_{X_2}(\phi_2 \mid \mathcal{B}_2) d\mu,$$

by the definition of the conditional expectation. If $B = B_1 \times B_2$ with $B_1 \in \mathcal{B}_1$, $B_2 \in \mathcal{B}_2$, then both sides of this equation are equal to

$$(119.4) \quad \left(\int_{B_1} \phi_1 d\mu_1 \right) \left(\int_{B_2} \phi_2 d\mu_2 \right),$$

using the definition of the conditional expectation again. This implies that the previous equation holds when B is the union of finitely many pairwise-disjoint sets of the form $B_1 \times B_2$, with $B_1 \in \mathcal{B}_1$ and $B_2 \in \mathcal{B}_2$. The analogous statement for any $B \in \mathcal{B}$ follows by approximation. If \mathcal{B}_1 or \mathcal{B}_2 is generated by a partition of X_1 or X_2 into finitely or countably many measurable sets, then every $B \in \mathcal{B}$ can be expressed as the union of finitely or countably many disjoint sets of the form $B_1 \times B_2$, with $B_1 \in \mathcal{B}_1$ and $B_2 \in \mathcal{B}_2$, as in Section 101, and the approximation is much simpler.

120 σ -Subalgebras and vectors

Let (X, \mathcal{A}, μ) be a probability space, and let \mathcal{B} be a σ -subalgebra of \mathcal{A} . Also let V be a finite-dimensional real or complex vector space with a norm, which can be identified with \mathbf{R}^n or \mathbf{C}^n for some n using a basis. Thus a V -valued function $f(x)$ on X corresponds an n -tuple $(f_1(x), \dots, f_n(x))$ of real or complex-valued functions on X . Such a function is considered to be integrable when its components $f_1(x), \dots, f_n(x)$ are integrable, in which case the integral is defined by integrating the components separately. Similarly, the conditional expectation of a V -valued function f on X may be defined by applying the conditional expectation to the components of f .

Let λ be a linear functional on V , so that $\lambda(v)$ can be expressed by a linear combination of the components of v . If $f(x)$ is an integrable V -valued function on X , then $\lambda(f(x))$ is an integrable real or complex-valued function on X , and

$$(120.1) \quad \lambda\left(\int_X f(x) d\mu(x)\right) = \int_X \lambda(f(x)) d\mu(x).$$

If $\|v\|$ is a norm on V , and $\|\lambda\|_*$ is the corresponding dual norm on V^* , then it follows that

$$(120.2) \quad \left| \lambda \left(\int_X f(x) d\mu(x) \right) \right| \leq \int_X |\lambda(f(x))| d\mu(x) \\ \leq \|\lambda\|_* \int_X \|f(x)\| d\mu(x).$$

This implies that

$$(120.3) \quad \left\| \int_X f(x) d\mu(x) \right\| \leq \int_X \|f(x)\| d\mu(x),$$

by the Hahn–Banach theorem. The same conclusion could also be obtained by approximating the integral by finite sums.

Similarly,

$$(120.4) \quad \lambda(E(f \mid \mathcal{B})) = E(\lambda(f) \mid \mathcal{B}),$$

and hence

$$(120.5) \quad |\lambda(E(f \mid \mathcal{B}))| \leq E(|\lambda(f)| \mid \mathcal{B}) \leq \|\lambda\|_* E(\|f\| \mid \mathcal{B}).$$

This implies that

$$(120.6) \quad \|E(f \mid \mathcal{B})\| \leq E(\|f\| \mid \mathcal{B}).$$

More precisely, if (120.5) holds at some point $x \in X$ for every linear functional λ on V , then (120.6) also holds at x , by the Hahn–Banach theorem. This works as well when (120.5) holds for every λ in a dense subset of

$$(120.7) \quad \{\lambda \in V^* : \|\lambda\|_* = 1\}.$$

Because V and hence V^* are finite-dimensional, there is a countable dense set in (120.7). If (120.5) holds almost everywhere on X for each $\lambda \in V^*$, then it holds simultaneously for a countable set of λ 's almost everywhere on X . This implies that (120.6) holds almost everywhere on X , as desired.

121 Martingales and products

Let (X, \mathcal{A}, μ) , (Y, \mathcal{B}, ν) be probability spaces, and suppose that their Cartesian product $X \times Y$ is equipped with the product probability measure $\mu \times \nu$. Also let $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \dots$ be an increasing sequence of σ -subalgebras of \mathcal{A} , and let $\widehat{\mathcal{B}}_j$ be the σ -subalgebra of the σ -algebra of measurable subsets of $X \times Y$ that corresponds to \mathcal{B}_j on X and \mathcal{B} on Y in the product space. As before, a function $F(x, y) \in L^p(X \times Y)$, $1 \leq p < \infty$, may be considered as representing an L^p function on X with values in $L^p(Y)$, and thus a martingale on $X \times Y$ with respect to $\widehat{\mathcal{B}}_j$ may be considered as representing a type of vector-valued martingale on X with respect to \mathcal{B}_j .

If $F(x, y) \in L^1(X \times Y)$, then put

$$(121.1) \quad I_Y(F)(x) = \int_Y F(x, y) d\nu(y).$$

Let us check that

$$(121.2) \quad I_Y(E_{X \times Y}(F \mid \widehat{\mathcal{B}}_j)) = E_X(I_Y(F) \mid \mathcal{B}_j)$$

for each j , where the subscripts of E indicate the spaces on which the conditional expectations are taken. Both sides of the equation are measurable functions on X with respect to \mathcal{B}_j , and so it is enough to show that

$$(121.3) \quad \int_A I_Y(E_{X \times Y}(F \mid \widehat{\mathcal{B}}_j)) d\mu = \int_A E_X(I_Y(F) \mid \mathcal{B}_j) d\mu$$

for every $A \in \mathcal{B}_j$. Of course,

$$(121.4) \quad \begin{aligned} \int_A I_Y(E_{X \times Y}(F \mid \widehat{\mathcal{B}}_j)) d\mu &= \int_{A \times Y} E_{X \times Y}(F \mid \widehat{\mathcal{B}}_j) d(\mu \times \nu) \\ &= \int_{A \times Y} F d(\mu \times \nu), \end{aligned}$$

because $A \times Y \in \widehat{\mathcal{B}}_j$. Similarly,

$$(121.5) \quad \int_A E_X(I_Y(F) \mid \mathcal{B}_j) d\mu = \int_A I_Y(F) d\mu = \int_{A \times Y} F d(\mu \times \nu).$$

Let us say that a measurable function $F(x, y) \in L^1(X \times Y)$ is *nice* if there are finitely many pairwise-disjoint measurable subsets B_1, \dots, B_n of Y with positive measure such that $F(x, y)$ is constant in y on B_k for $k = 1, \dots, n$. If $\phi_k(x) = F(x, y)$ when $y \in B_k$, then $\phi_k(x) \in L^1(X)$ for each k , and

$$(121.6) \quad F(x, y) = \sum_{k=1}^n \phi_k(x) \mathbf{1}_{B_k}(y),$$

where $\mathbf{1}_{B_k}(y)$ is the indicator function associated to B_k on Y , equal to 1 when $y \in B_k$ and to 0 when $y \in Y \setminus B_k$. In this case,

$$(121.7) \quad E_{X \times Y}(F \mid \widehat{\mathcal{B}}_j)(x, y) = \sum_{k=1}^n E_X(\phi_k \mid \mathcal{B}_j)(x) \mathbf{1}_{B_k}(y),$$

as in Section 119. In effect, F corresponds to a function on X with values in an n -dimensional vector space under these conditions.

Suppose that $F(x, y) \in L^p(X \times Y)$, $1 \leq p < \infty$, and put

$$(121.8) \quad N_p(F)(x) = \left(\int_Y |F(x, y)|^p d\nu(y) \right)^{1/p}.$$

Thus $N_p(F) \in L^p(X)$, and

$$(121.9) \quad \left(\int_X N_p(F)(x)^p d\mu(x) \right)^{1/p} = \left(\int_{X \times Y} |F(x, y)|^p d(\mu \times \nu)(x, y) \right)^{1/p}.$$

We would like to check that

$$(121.10) \quad N_p(E_{X \times Y}(F | \widehat{\mathcal{B}}_j))(x) \leq E_X(N_p(F) | \mathcal{B}_j)(x)$$

for almost every $x \in X$ and each $j \geq 1$. If $p = 1$, then $N_1(F) = I_Y(|F|)$, and

$$(121.11) \quad \begin{aligned} I_Y(|E_{X \times Y}(F | \widehat{\mathcal{B}}_j)|) &\leq I_Y(E_{X \times Y}(|F| | \widehat{\mathcal{B}}_j)) \\ &= E_X(I_Y(|F|) | \mathcal{B}_j). \end{aligned}$$

If $p > 1$ and F is nice, then (121.10) follows from the discussion in the preceding section. More precisely, one can take V to be the n -dimensional vector space spanned by $\mathbf{1}_{B_1}, \dots, \mathbf{1}_{B_n}$, equipped with the $L^p(Y)$ norm. Otherwise, one can approximate F by nice functions in $L^p(X \times Y)$.

If $\{F_j\}_{j=1}^\infty$ is a martingale on $X \times Y$ with respect to the $\widehat{\mathcal{B}}_j$'s such that $F_j \in L^p(X \times Y)$ for each j , $1 \leq p < \infty$, then it follows that $\{N_p(F_j)\}_{j=1}^\infty$ is a submartingale on X with respect to the \mathcal{B}_j 's. This leads to the same type of maximal function estimates as before. If $\{F_j\}_{j=1}^\infty$ converges in $L^p(X \times Y)$, then one may conclude that $\{F_j(x, \cdot)\}_{j=1}^\infty$ converges in $L^p(Y)$ for almost every $x \in X$. In particular, this holds when $1 < p < \infty$ and the norm of $F_j(x, y)$ in $L^p(X \times Y)$ is uniformly bounded in j .

Instead of (121.10), it is easier to show that

$$(121.12) \quad I_Y(|E_{X \times Y}(F | \widehat{\mathcal{B}}_j)|^p) \leq E_X(I_Y(|F|^p) | \mathcal{B}_j)$$

almost everywhere on X . As in the $p = 1$ case, one has that

$$(121.13) \quad \begin{aligned} I_Y(|E_{X \times Y}(F | \widehat{\mathcal{B}}_j)|^p) &\leq I_Y(E_{X \times Y}(|F|^p | \widehat{\mathcal{B}}_j)) \\ &= E_X(I_Y(|F|^p) | \mathcal{B}_j) \end{aligned}$$

when $p > 1$. If $\{F_j\}_{j=1}^\infty$ is a martingale on $X \times Y$ with respect to the $\widehat{\mathcal{B}}_j$'s such that $F_j \in L^p(X \times Y)$ for each j , then this implies the less precise statement that $I_Y(|F_j|^p) = N_p(F_j)^p$ is a submartingale on X with respect to the \mathcal{B}_j 's. One can still get some maximal function estimates from this, which are adequate for the same conclusions about pointwise convergence.

If Y' is a σ -finite measure space, then one can choose a positive weight on Y' to get a probability measure, as in Section 109. This permits one to identify $L^p(Y')$ with $L^p(Y)$ for a probability space Y , as before. Thus martingales on X with values in $L^p(Y')$ can be identified with martingales on X with values in $L^p(Y)$, to which the discussion in this section applies.

122 ℓ^p -Valued martingales

Let (X, \mathcal{A}, μ) be a probability space, and let $\{f_l(x)\}_{l=1}^\infty$ be a sequence of real or complex-valued functions on X such that $f_l(x) \in L^p(X)$ for each l , $1 \leq p < \infty$, and

$$(122.1) \quad \sum_{l=1}^\infty \int_X |f_l(x)|^p d\mu(x) < \infty.$$

This is the same as

$$(122.2) \quad \int_X \sum_{l=1}^{\infty} |f_l(x)|^p d\mu(x) < \infty,$$

which implies that $\{f_l(x)\}_{l=1}^{\infty} \in \ell^p = \ell^p(\mathbf{Z}_+)$ for almost every $x \in X$. One can also think of $\{f_l(x)\}_{l=1}^{\infty}$ as an element of $L^p(X \times \mathbf{Z}_+)$, where $X \times \mathbf{Z}_+$ is equipped with the product measure associated to counting measure on \mathbf{Z}_+ .

If \mathcal{B} is a σ -subalgebra of \mathcal{A} , then of course one can take the conditional expectation $E(f_l | \mathcal{B})$ of f_l for each l , and

$$(122.3) \quad \int_X |E(f_l | \mathcal{B})|^p d\mu(x) \leq \int_X E(|f_l|^p | \mathcal{B}) d\mu(x) = \int_X |f_l|^p d\mu.$$

Hence

$$(122.4) \quad \sum_{l=1}^{\infty} \int_X |E(f_l | \mathcal{B})|^p d\mu \leq \sum_{l=1}^{\infty} \int_X |f_l|^p d\mu.$$

This is another way to look at conditional expectation of ℓ^p -valued functions, which is consistent with the earlier discussions.

More precisely,

$$(122.5) \quad |E(f_l | \mathcal{B})|^p \leq E(|f_l|^p | \mathcal{B})$$

almost everywhere on X for each l , and so

$$(122.6) \quad \sum_{l=1}^{\infty} |E(f_l | \mathcal{B})|^p \leq \sum_{l=1}^{\infty} E(|f_l|^p | \mathcal{B}) = E\left(\sum_{l=1}^{\infty} |f_l|^p | \mathcal{B}\right)$$

almost everywhere on X . As in Section 120,

$$(122.7) \quad \left(\sum_{l=1}^n |E(f_l | \mathcal{B})|^p\right)^{1/p} \leq E\left(\left(\sum_{l=1}^n |f_l|^p\right)^{1/p} | \mathcal{B}\right)$$

almost everywhere on X for each $n \in \mathbf{Z}_+$. This implies that

$$(122.8) \quad \left(\sum_{l=1}^n |E(f_l | \mathcal{B})|^p\right)^{1/p} \leq E\left(\left(\sum_{l=1}^{\infty} |f_l|^p\right)^{1/p} | \mathcal{B}\right)$$

almost everywhere on X for each n , and thus

$$(122.9) \quad \left(\sum_{l=1}^{\infty} |E(f_l | \mathcal{B})|^p\right)^{1/p} \leq E\left(\left(\sum_{l=1}^{\infty} |f_l|^p\right)^{1/p} | \mathcal{B}\right).$$

As in Section 109, one can choose a positive weight on \mathbf{Z}_+ to identify ℓ^p with $L^p(Y)$, where Y is a probability space. Thus the estimates in the preceding paragraph can be seen as a special case of those in the previous section, with simplifications from the discreteness of Y . As before, one can get submartingales from the norms of ℓ^p -valued martingales, and then maximal function estimates for these. In particular, it follows that an ℓ^1 -valued martingale with bounded L^1 norm converges almost everywhere, as in Section 111.

123 Approximation in product spaces

Let $(X_1, \mathcal{A}_1, \mu_1)$, $(X_2, \mathcal{A}_2, \mu_2)$ be measure spaces with $\mu_1(X_1), \mu_2(X_2) < \infty$, and consider their Cartesian product $X_1 \times X_2$. The σ -algebra \mathcal{A} of measurable subsets of X is defined as the smallest σ -algebra of subsets of X that contains the measurable rectangles $A_1 \times A_2$, $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$. Note that the intersection of two measurable rectangles in X is also a measurable rectangle, and that the complement of a measurable rectangle is the union of three pairwise-disjoint measurable rectangles, since

$$(123.1) \quad (X_1 \times X_2) \setminus (A_1 \times A_2) = ((X_1 \setminus A_1) \times A_2) \cup (A_1 \times (X_2 \setminus A_2)) \cup ((X_1 \setminus A_1) \times (X_2 \setminus A_2)).$$

Let \mathcal{E} be the collection of subsets of X that can be expressed as the union of finitely many pairwise-disjoint measurable rectangles. This is an algebra of subsets of X , by the previous observations. Also let $\mu = \mu_1 \times \mu_2$ be the product measure associated to μ_1, μ_2 on \mathcal{A} . If

$$(123.2) \quad d(A, B) = \mu(A \triangle B)$$

is the corresponding semimetric on \mathcal{A} as in Section 79, then \mathcal{E} is dense in \mathcal{A} with respect to $d(A, B)$. Depending on the way that the product measure is defined, this may be obvious from the construction. At any rate, this follows from the discussion in Section 79, which implies that the closure $\overline{\mathcal{E}}$ of \mathcal{E} in \mathcal{A} is a σ -subalgebra of \mathcal{A} that contains \mathcal{E} . One could also use the characterization of \mathcal{A} as the smallest monotone class of subsets of X that contains \mathcal{E} . If X_1, X_2 are σ -finite and $A \subseteq X$ is a measurable set with $\mu(A) < \infty$, then one can first approximate A by subsets of products of measurable sets with finite measure, and then continue as before to approximate A by elements of \mathcal{E} . Using these approximations, one can check that nice functions are dense in $L^p(X)$ when $1 \leq p < \infty$, as in Section 121. Of course, these statements are much simpler when X_1 or X_2 has only finitely or countably many elements and all of its subsets are measurable, or when \mathcal{A}_1 or \mathcal{A}_2 is generated by a partition of the corresponding space into finitely or countably many subsets.

124 Mixed norms

Let (X, \mathcal{A}, μ) , (Y, \mathcal{B}, ν) be probability spaces, and let their Cartesian product $X \times Y$ be equipped with the product measure $\mu \times \nu$, as usual. Consider the space of real or complex-valued measurable functions $F(x, y)$ on $X \times Y$ such that

$$(124.1) \quad \int_X \left(\int_Y |F(x, y)|^p d\nu(y) \right)^{1/p} d\mu(x)$$

is finite, where $1 \leq p < \infty$. It is easy to see that this is a vector space, and that (124.1) becomes a norm on this space when we identify functions that are equal

almost everywhere. If $F(x, y) \in L^p(X \times Y)$, then $F(x, y)$ is in this space, and

$$(124.2) \quad \int_X \left(\int_Y |F(x, y)|^p d\nu(y) \right)^{1/p} d\mu(x) \\ \leq \left(\int_{X \times Y} |F(x, y)|^p d(\mu \times \nu)(x, y) \right)^{1/p},$$

by Fubini's theorem and Jensen's inequality. Similarly, if $F(x, y)$ is in this space, then $F(x, y) \in L^1(X \times Y)$, and

$$(124.3) \quad \int_{X \times Y} |F(x, y)| d(\mu \times \nu)(x, y) \\ \leq \int_X \left(\int_Y |F(x, y)|^p d\nu(y) \right)^{1/p} d\mu(x),$$

again by Fubini's theorem and Jensen's inequality.

Suppose that $F(x, y)$ is in this space, and put

$$(124.4) \quad N_p(F)(x) = \left(\int_Y |F(x, y)|^p d\nu(y) \right)^{1/p},$$

as in Section 121. Thus (124.1) is the same as the $L^1(X)$ norm of $N_p(F)(x)$. If $L \geq 0$, then define $F_L(x, y)$ on $X \times Y$ by

$$(124.5) \quad \begin{aligned} F_L(x, y) &= F(x, y) \text{ when } N_p(F)(x) \leq L, \\ &= 0 \quad \text{when } N_p(F)(x) > L. \end{aligned}$$

In particular, $N_p(F_L)(x) = N_p(F)(x)$ when $N_p(F)(x) \leq L$, and $N_p(F_L)(x) = 0$ when $N_p(F)(x) > L$. It follows that $F_L(x, y) \in L^p(X \times Y)$ for each L , and that $F_L(x, y)$ converges to $F(x, y)$ with respect to the norm (124.1) as $L \rightarrow \infty$, so that $L^p(X \times Y)$ is a dense linear subspace of this space.

Let $\{F_j(x, y)\}_{j=1}^\infty$ be a sequence of measurable functions on $X \times Y$. By Fatou's lemma,

$$(124.6) \quad \int_Y \liminf_{j \rightarrow \infty} |F_j(x, y)|^p d\nu(y) \leq \liminf_{j \rightarrow \infty} \int_Y |F_j(x, y)|^p d\nu(y)$$

for every $x \in X$. Equivalently,

$$(124.7) \quad \left(\int_Y \left(\liminf_{j \rightarrow \infty} |F_j(x, y)| \right)^p d\nu(y) \right)^{1/p} \\ \leq \liminf_{j \rightarrow \infty} \left(\int_Y |F_j(x, y)|^p d\nu(y) \right)^{1/p}.$$

for each $x \in X$. Applying Fatou's lemma a second time, we get that

$$(124.8) \quad \int_X \left(\int_Y \left(\liminf_{j \rightarrow \infty} |F_j(x, y)| \right)^p d\nu(y) \right)^{1/p} d\mu(x) \\ \leq \liminf_{j \rightarrow \infty} \int_X \left(\int_Y |F_j(x, y)|^p d\nu(y) \right)^{1/p} d\mu(x).$$

Suppose that $\{F_j(x, y)\}_{j=1}^\infty$ converges almost everywhere to $F(x, y)$ on $X \times Y$. It follows that for almost every $x \in X$, $\{F_j(x, y)\}_{j=1}^\infty$ converges to $F(x, y)$ for almost every $y \in Y$. Hence

$$(124.9) \quad \int_Y |F(x, y)|^p d\nu(y) \leq \liminf_{j \rightarrow \infty} \int_Y |F_j(x, y)|^p d\nu(y)$$

for almost every $x \in X$. This implies that

$$(124.10) \quad \begin{aligned} \int_X \left(\int_Y |F(x, y)|^p d\nu(y) \right)^{1/p} d\mu(x) \\ \leq \liminf_{j \rightarrow \infty} \int_X \left(\int_Y |F_j(x, y)|^p d\nu(y) \right)^{1/p} d\mu(x), \end{aligned}$$

as before.

125 Mixed-norm martingales

Let (X, \mathcal{A}, μ) , (Y, \mathcal{B}, ν) be probability spaces, and let $X \times Y$ be equipped with the product measure $\mu \times \nu$. Also let $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \dots$ be an increasing sequence of σ -subalgebras of \mathcal{A} , and let $\widehat{\mathcal{B}}_j$ be the σ -algebra of subsets of $X \times Y$ that corresponds to \mathcal{B}_j on X and \mathcal{B} on Y in the product space. Suppose that $F(x, y)$ is a measurable function on $X \times Y$ for which (124.1) is finite, $1 \leq p < \infty$. In particular, $F(x, y) \in L^1(X \times Y)$, and so

$$(125.1) \quad F_j = E(F | \widehat{\mathcal{B}}_j)$$

defines a martingale on $X \times Y$ with respect to $\widehat{\mathcal{B}}_j$.

As in Section 121,

$$(125.2) \quad N_p(F_j) \leq E_X(N_p(F) | \mathcal{B}_j)$$

almost everywhere on X for each $j \geq 1$, where the subscript X of E indicates that the conditional expectation is taken on X . More precisely, this is the same as (121.10) when $F(x, y) \in L^p(X \times Y)$, and otherwise we can approximate $F(x, y)$ by elements of $L^p(X \times Y)$ with respect to the norm (124.1), as in the previous section. Integrating (125.2) over X , we get that

$$(125.3) \quad \int_X N_p(F_j) d\mu \leq \int_X E_X(N_p(F) | \mathcal{B}_j) d\mu = \int_X N_p(F) d\mu$$

for each j . Thus the norm of F_j with respect to (124.1) is less than or equal to (124.1) for each j .

One can also check that $\{F_j\}_{j=1}^\infty$ converges to F with respect to the norm (124.1). If $F \in L^p(X \times Y)$, then $\{F_j\}_{j=1}^\infty$ converges to F with respect to the L^p norm, and hence with respect to (124.1). Otherwise, one can approximate F by elements of $L^p(X \times Y)$, using the uniform bound for the norm of F_j in the previous paragraph.

If we apply (125.2) to F_{j+1} instead of F_j , then we get that

$$(125.4) \quad N_p(F_j) \leq E_X(N_p(F_{j+1}) \mid \mathcal{B}_j)$$

almost everywhere on X for each $j \geq 1$. Thus $\{N_p(F_j)\}_{j=1}^\infty$ is a submartingale on X with respect to the \mathcal{B}_j 's, which leads to maximal function estimates as before. Using convergence of $\{F_j\}_{j=1}^\infty$ to F with respect to the norm (124.1), one can show that $\{F_j(x, \cdot)\}_{j=1}^\infty$ converges to $F(x, \cdot)$ in $L^p(Y)$ for almost every $x \in X$. This is basically the same as in the previous situations, once we have the same ingredients as before.

126 Mixed-norm convergence

Let (X, \mathcal{A}, μ) , (Y, \mathcal{B}, ν) be probability spaces, and let $X \times Y$ be equipped with $\mu \times \nu$, as usual. Also let $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \dots$ be an increasing sequence of σ -subalgebras of \mathcal{A} , and let $\widehat{\mathcal{B}}_j$ be the σ -algebra of subsets of $X \times Y$ that corresponds to \mathcal{B}_j on X and \mathcal{B} on Y in the product space. Suppose that $\{F_j\}_{j=1}^\infty$ is a martingale on $X \times Y$ with respect to the $\widehat{\mathcal{B}}_j$'s whose norms as in (124.1) are uniformly bounded for some $p > 1$. Equivalently,

$$(126.1) \quad \int_X N_p(F_j) d\mu \leq C$$

for some $C \geq 0$ and each j . Note that $\{N_p(F_j)\}_{j=1}^\infty$ is a submartingale on X with respect to the \mathcal{B}_j 's, as in (125.4).

Suppose in addition that $\{N_p(F_j)\}_{j=1}^\infty$ is uniformly integrable on X , as in Section 83, and let us check that $\{F_j\}_{j=1}^\infty$ is uniformly integrable on $X \times Y$. Let $\epsilon > 0$ be given, and choose $\delta > 0$ such that

$$(126.2) \quad \int_A N_p(F_j) d\mu < \frac{\epsilon}{2}$$

for every measurable set $A \subseteq X$ with $\mu(A) < \delta$ and each j . If

$$(126.3) \quad A_{j,L} = \{x \in X : N_p(F_j)(x) > L\},$$

then

$$(126.4) \quad \mu(A_{j,L}) < L^{-1} C$$

for each j , L , by Tchebychev's inequality. Hence $\mu(A_{j,L}) < \delta$ for each j when L is sufficiently large, which implies that

$$(126.5) \quad \int_{A_{j,L} \times Y} |F_j| d(\mu \times \nu) \leq \int_{A_{j,L}} N_p(F_j) d\mu < \frac{\epsilon}{2}$$

for each j when L is sufficiently large. On the complement of $A_{j,L} \times Y$, we have that

$$(126.6) \quad \int_{(X \setminus A_{j,L}) \times Y} |F_j|^p d(\mu \times \nu) = \int_{X \setminus A_{j,L}} N_p(F_j)^p d\mu \leq L^p$$

for each j , L , by the definition of $A_{j,L}$. Let q be the exponent conjugate to p , so that $1/p + 1/q = 1$. If $B \subseteq X \times Y$ is measurable, then

$$(126.7) \quad \int_B |F_j| d(\mu \times \nu) \leq ((\mu \times \nu)(B))^{1/q} \left(\int_B |F_j|^p d(\mu \times \nu) \right)^{1/p},$$

by Hölder's inequality. If $B \subseteq (X \setminus A_{j,L}) \times Y$, then it follows that

$$(126.8) \quad \int_B |F_j| d(\mu \times \nu) \leq L ((\mu \times \nu)(B))^{1/q}.$$

In order to show that $\{F_j\}_{j=1}^\infty$ is uniformly integrable, one can combine this with the earlier estimate for the integral of $|F_j|$ over $A_{j,L} \times Y$ when L is sufficiently large.

If $\{F_j\}_{j=1}^\infty$ is uniformly integrable on $X \times Y$, then $\{F_j\}_{j=1}^\infty$ converges in $L^1(X \times Y)$ to a function F , and $F_j = E(F | \widehat{\mathcal{B}}_j)$ for each j . Moreover, $\{F_j\}_{j=1}^\infty$ converges to F almost everywhere on $X \times Y$, which implies that the norm of F with respect to (124.1) is also finite, as in Section 124. Thus we are back in the situation of the preceding section. This implies that $\{F_j\}_{j=1}^\infty$ also converges to F with respect to the norm (124.1), and that $\{F_j(x, \cdot)\}_{j=1}^\infty$ converges to $F(x, \cdot)$ in $L^p(Y)$ for almost every $x \in X$.

Suppose now that $\{N_p(F_j)\}_{j=1}^\infty$ is still bounded in $L^1(X)$, but may not be uniformly integrable. Because $\{N_p(F_j)\}_{j=1}^\infty$ is a submartingale with respect to the \mathcal{B}_j 's, the corresponding maximal function can be estimated in the usual way. In this case, $\{F_j\}_{j=1}^\infty$ can be approximated by martingales $\{G_j\}_{j=1}^\infty$ on $X \times Y$ such that $\{N_p(G_j)\}_{j=1}^\infty$ is uniformly integrable, as in Section 85. More precisely, the approximation basically takes place in the x variable. This permits one to show that $\{F_j(x, \cdot)\}_{j=1}^\infty$ converges in $L^p(Y)$ for almost every $x \in X$, as before.

127 The ℓ^p version

Let (X, \mathcal{A}, μ) be a probability space, and let $1 \leq p < \infty$ be given. If $\{f_l(x)\}_{l=1}^\infty$ is a sequence of real or complex-valued measurable functions on X such that

$$(127.1) \quad \int_X \left(\sum_{l=1}^\infty |f_l(x)|^p \right)^{1/p} d\mu(x)$$

is finite, then

$$(127.2) \quad \sum_{l=1}^\infty |f_l(x)|^p < \infty$$

for almost every $x \in X$. It is easy to see that the space of sequences of functions on X of this type is a vector space, and that (127.1) defines a norm on this vector space when we identify functions that are equal almost everywhere on X . We can also use a weight on the set of positive integers to identify ℓ^p with $L^p(Y)$ for a probability space Y , so that this expression is the same as (124.1).

If $\{f_l(x)\}_{l=1}^\infty$ is a sequence of functions in $L^p(X)$ such that

$$(127.3) \quad \sum_{l=1}^\infty \int_X |f_l(x)|^p d\mu(x) = \int_X \sum_{l=1}^\infty |f_l(x)|^p d\mu(x) < \infty,$$

then (127.1) is also finite, because

$$(127.4) \quad \int_X \left(\sum_{l=1}^\infty |f_l(x)|^p \right)^{1/p} d\mu(x) \leq \left(\int_X \sum_{l=1}^\infty |f_l(x)|^p d\mu(x) \right)^{1/p},$$

by Jensen's inequality. These sequences of functions are dense among those for which (127.1) is finite, with respect to the norm (127.1), for the same reasons as in Section 124. Of course, these two conditions on sequences of functions on X are the same when $p = 1$. Alternatively, if $\{f_l(x)\}_{l=1}^\infty$ is a sequence of functions on X for which (127.1) is finite, then

$$(127.5) \quad \lim_{n \rightarrow \infty} \int_X \left(\sum_{l=n}^\infty |f_l(x)|^p \right)^{1/p} d\mu(x) = 0,$$

by the dominated convergence theorem. This implies that $\{f_l(x)\}_{l=1}^\infty$ can be approximated by sequences of functions for which all but finitely many terms are equal to 0 with respect to the norm (127.1).

Let $f(x) = \{f_l(x)\}_{l=1}^\infty$ be a sequence of measurable functions on X for which (127.1) is finite, and let \mathcal{B} be a σ -subalgebra of \mathcal{A} . As in Sections 120 and 122,

$$(127.6) \quad \left(\sum_{l=1}^n |E(f_l | \mathcal{B})|^p \right)^{1/p} \leq E \left(\left(\sum_{l=1}^n |f_l|^p \right)^{1/p} | \mathcal{B} \right)$$

almost everywhere on X for each n , and hence

$$(127.7) \quad \left(\sum_{l=1}^\infty |E(f_l | \mathcal{B})|^p \right)^{1/p} \leq E \left(\left(\sum_{l=1}^\infty |f_l|^p \right)^{1/p} | \mathcal{B} \right)$$

almost everywhere on X . In particular,

$$(127.8) \quad \begin{aligned} \int_X \left(\sum_{l=1}^\infty |E(f_l | \mathcal{B})|^p \right)^{1/p} d\mu &\leq \int_X E \left(\left(\sum_{l=1}^\infty |f_l|^p \right)^{1/p} | \mathcal{B} \right) d\mu \\ &= \int_X \left(\sum_{l=1}^\infty |f_l|^p \right)^{1/p} d\mu. \end{aligned}$$

Now let $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \dots$ be an increasing sequence of σ -subalgebras of \mathcal{A} . Also let $f_j(x) = \{f_{j,l}(x)\}_{l=1}^\infty$ be a sequence of measurable functions with respect to \mathcal{B}_j for which (127.1) is finite for each j , and put

$$(127.9) \quad A_j = \int_X \left(\sum_{l=1}^\infty |f_{j,l}(x)|^p \right)^{1/p} d\mu(x).$$

Suppose that $\{f_{j,l}\}_{l=1}^\infty$ is a martingale with respect to this filtration for each j , so that $f_{j,l} = E(f_{j+1,l} \mid \mathcal{B}_j)$ for each $j, l \geq 1$. Thus

$$(127.10) \quad \|f_j(x)\|_p = \left(\sum_{l=1}^\infty |f_{j,l}(x)|^p \right)^{1/p}$$

is a submartingale with respect to this filtration, as in the previous paragraph. This implies that $\{A_j\}_{j=1}^\infty$ is monotone increasing, as usual. Similarly, if

$$(127.11) \quad A_{j,n} = \int_X \left(\sum_{l=1}^n |f_{j,l}(x)|^p \right)^{1/p} d\mu(x),$$

then $A_{j,n} \leq A_{j+1,n}$ for each $j, n \geq 1$. Note that

$$(127.12) \quad \lim_{n \rightarrow \infty} A_{j,n} = A_j$$

for each j , by the dominated convergence theorem.

Suppose that the A_j 's are bounded, and put

$$(127.13) \quad A = \sup_{j \geq 1} A_j.$$

Let $\delta > 0$ be given, and choose j_0 such that

$$(127.14) \quad A_{j_0} > A - \delta.$$

Because $A_{j_0,n} \rightarrow A_{j_0}$ as $n \rightarrow \infty$, we can choose n_0 so that

$$(127.15) \quad A_{j_0,n_0} > A - \delta.$$

If $j \geq j_0$, then monotonicity implies that

$$(127.16) \quad A_{j,n_0} > A - \delta.$$

Let us pause a moment to record some elementary inequalities that will be helpful later. If $a, b > 0$, then

$$(127.17) \quad (a+b)^{1/p} \geq a^{1/p} + p^{-1} (a+b)^{(1/p)-1} b.$$

This follows from calculus, because

$$(127.18) \quad \frac{d}{dt} (a+t)^{1/p} = p^{-1} (a+t)^{(1/p)-1}$$

is minimized on $[0, b]$ at $t = b$. Remember that $0 < 1/p \leq 1$, because $1 \leq p < \infty$. If $b \geq \epsilon a$ for some $\epsilon > 0$, then $a+b \leq (\epsilon^{-1} + 1)b$, and so

$$(127.19) \quad (a+b)^{1/p} \geq a^{1/p} + p^{-1} (\epsilon^{-1} + 1)^{(1/p)-1} b^{1/p}.$$

This implies that

$$(127.20) \quad b^{1/p} \leq \epsilon^{1/p} a^{1/p} + p(\epsilon^{-1} + 1)^{1-(1/p)} ((a+b)^{1/p} - a^{1/p}),$$

for every $\epsilon > 0$. More precisely, $b^{1/p}$ is less than or equal to the second term on the right when $b \geq \epsilon a$, by the previous inequality, and otherwise $b^{1/p}$ is less than or equal to the first term on the right, because $b < \epsilon a$. Note that (127.20) also holds when $a = 0$ or $b = 0$.

Let us apply (127.20) to

$$(127.21) \quad a = \sum_{l=1}^n |f_{j,l}(x)|^p, \quad b = \sum_{l=n+1}^{\infty} |f_{j,l}(x)|^p,$$

using also the fact that $a^{1/p} \leq (a+b)^{1/p} = \|f_j(x)\|_p$. This implies that

$$(127.22) \quad \left(\sum_{l=n+1}^{\infty} |f_{j,l}(x)|^p \right)^{1/p} \leq \epsilon^{1/p} \|f_j(x)\|_p + p(\epsilon^{-1} + 1)^{1-(1/p)} \left(\|f_j(x)\|_p - \left(\sum_{l=1}^n |f_{j,l}(x)|^p \right)^{1/p} \right).$$

Integrating over X , we get that

$$(127.23) \quad \int_X \left(\sum_{l=n+1}^{\infty} |f_{j,l}(x)|^p \right)^{1/p} d\mu(x) \leq \epsilon^{1/p} A + p(\epsilon^{-1} + 1)^{1-(1/p)} (A - A_{j,n}),$$

using also the fact that $A_j \leq A$ for each j , by the definition of A . Taking $n = n_0$, we get that

$$(127.24) \quad \int_X \left(\sum_{l=n_0+1}^{\infty} |f_{j,l}(x)|^p \right)^{1/p} d\mu(x) \leq \epsilon^{1/p} A + p(\epsilon^{-1} + 1)^{1-(1/p)} \delta$$

when $j \geq j_0$.

If $\eta > 0$ is given, then we can first choose $\epsilon > 0$ so that $\epsilon^{1/p} A < \eta/2$, and then choose δ depending on ϵ such that $p(\epsilon^{-1} + 1)^{1-(1/p)} \delta < \eta/2$. If j_0, n_0 are as before, then (127.24) implies that

$$(127.25) \quad \int_X \left(\sum_{l=n_0+1}^{\infty} |f_{j,l}(x)|^p \right)^{1/p} d\mu(x) < \eta$$

when $j \geq j_0$. The integral on the left side of (127.25) is actually monotone increasing in j , for the usual submartingale reasons, which implies that (127.25) holds for every j . Put $g_{j,l}(x) = f_{j,l}(x)$ when $l \leq n_0$ and $g_{j,l}(x) = 0$ when $l > n_0$, so that $\{g_{j,l}\}_{l=1}^{\infty}$ is a martingale with respect to the \mathcal{B}_j 's for each l , and $f_j(x) = \{f_{j,l}(x)\}_{l=1}^{\infty}$ is approximated by $g_j(x) = \{g_{j,l}(x)\}_{l=1}^{\infty}$ uniformly in j with respect to the norm (127.1), by (127.25). Using this approximation and maximal function estimates for $\|f_j(x) - g_j(x)\|_p$, one can show that $\{f_j(x)\}_{j=1}^{\infty}$ converges in ℓ^p for almost every $x \in X$, as in Sections 110 and 111.

128 The doubling condition

Let (X, \mathcal{A}, μ) be a probability space, and let $\mathcal{P}_0, \mathcal{P}_1, \mathcal{P}_2, \dots$ be a sequence of partitions of X into finitely many measurable sets of positive measure such that \mathcal{P}_{j+1} is a refinement of \mathcal{P}_j for each j and \mathcal{P}_0 is the trivial partition consisting of only X itself. We say that the \mathcal{P}_j 's satisfy a doubling condition if there is a $C \geq 1$ such that

$$(128.1) \quad \mu(A) \leq C \mu(B)$$

when $A \in \mathcal{P}_j$, $B \in \mathcal{P}_{j+1}$, and $B \subseteq A$. This implies that for each $A \in \mathcal{P}_j$ there are less than or equal to C sets $B \in \mathcal{P}_{j+1}$ such that $B \subseteq A$. In particular, this implies that \mathcal{P}_j has less than or equal to C^j elements for each j . If $X = [0, 1)$ is equipped with Lebesgue measure and \mathcal{P}_j consists of the dyadic subintervals of $[0, 1)$ with length 2^{-j} , then (128.1) holds with $C = 2$.

Let $\mathcal{B}_j = \mathcal{B}(\mathcal{P}_j)$ be the σ -subalgebra of \mathcal{A} generated by \mathcal{P}_j , as in Section 77. Thus $\mathcal{B}_j \subseteq \mathcal{B}_{j+1}$ for each j , since \mathcal{P}_{j+1} is supposed to be a refinement of \mathcal{P}_j . If $f_{j+1}(x)$ is a nonnegative real-valued function on X which is measurable with respect to \mathcal{B}_{j+1} for some $j \geq 0$, then

$$(128.2) \quad f_{j+1} \leq C E(f_{j+1} \mid \mathcal{B}_j).$$

If $\{f_j\}_{j=0}^\infty$ is a martingale with respect to this filtration consisting of nonnegative real-valued functions, then

$$(128.3) \quad f_{j+1} \leq C f_j$$

for each j .

Suppose now that $\{\phi_j\}_{j=0}^\infty$ is a submartingale on X with respect to this filtration consisting of nonnegative real-valued functions, and put

$$(128.4) \quad \psi_j = E(\phi_{j+1} \mid \mathcal{B}_j).$$

Thus

$$(128.5) \quad \phi_j \leq \psi_j$$

for each $j \geq 0$, because $\{\phi_j\}_{j=1}^\infty$ is a submartingale. Hence

$$(128.6) \quad \psi_j = E(\phi_{j+1} \mid \mathcal{B}_j) \leq E(\psi_{j+1} \mid \mathcal{B}_j),$$

which implies that $\{\psi_j\}_{j=0}^\infty$ is also a submartingale. The doubling condition implies that

$$(128.7) \quad \phi_{j+1} \leq C \psi_j$$

for each $j \geq 0$, as in (128.2). If the ϕ_j 's have bounded L^p norm for some $p \geq 1$, then the ψ_j 's have bounded L^p norm as well, and with the same bound.

Let V be a real or complex vector space with a norm $\|v\|$, and let $\{f_j(x)\}_{j=0}^\infty$ be a V -valued martingale on X with respect to the \mathcal{B}_j 's, as in Section 103. Thus $\phi_j(x) = \|f_j(x)\|$ is a nonnegative real-valued submartingale on X , and $\psi_j(x)$ can be defined as in the previous paragraph. Note that $f_0(x)$ is constant on X , and let $t \geq \|f_0(x)\|$ be given. Put $\tau(x) = \infty$ when $\psi_j(x) \leq t$ for each $j \geq 0$, and

otherwise let $\tau(x)$ be the smallest nonnegative integer l such that $\psi_l(x) > t$. This is a stopping time, as in Section 96. If $\tau_n(x) = \min(\tau(x), n)$, then

$$(128.8) \quad g_n(x) = f_{\tau_n(x)}(x)$$

is also a V -valued martingale on X , as before. This is basically the same as the approximation to $\{f_j(x)\}_{j=0}^\infty$ described in Section 85, except that we use the maximal function associated to ψ_j instead of ϕ_j . By construction,

$$(128.9) \quad \|g_n(x)\| = \|f_n(x)\| \leq \psi_n(x) \leq t$$

when $n < \tau(x)$, and

$$(128.10) \quad \|g_n(x)\| = \|f_{\tau(x)}(x)\| \leq C \psi_{\tau(x)-1}(x) \leq C t$$

when $0 < \tau(x) \leq n$, because of the doubling condition. It follows that

$$(128.11) \quad \|g_n(x)\| \leq C t$$

for every $x \in X$ and $n \geq 0$, since $\|g_n(x)\| = \|f_0(x)\| \leq t$ when $\tau(x) = 0$, by hypothesis. This is analogous to (85.16), with the integrable function $h(x)$ replaced by Ct . If $\phi_j(x) = \|f_j(x)\|$ has bounded L^1 norm, so that $\psi_j(x)$ has bounded L^1 norm too, then the measure of the set where $\tau(x) < \infty$ can be estimated as before. Of course, $g_n(x) = f_n(x)$ for every $n \geq 0$ when $\tau(x) = \infty$. If every uniformly bounded V -valued martingale on X converges almost everywhere, then every V -valued martingale $\{f_j(x)\}_{j=1}^\infty$ such that $\|f_j(x)\|$ has bounded L^1 norm also converges almost everywhere, as in Section 85.

129 Paths and martingales

Let V be a real or complex vector space with a norm $\|v\|$, and let F be a V -valued function on $[0, 1]$. If $[a, b)$ is a dyadic subinterval of $[0, 1]$ of length $b - a = 2^{-j}$, then put

$$(129.1) \quad f_j(x) = 2^j (F(b) - F(a))$$

for every $x \in [a, b)$. This defines $f_j(x)$ as a V -valued function on $[0, 1]$ which is constant on the dyadic intervals of length 2^{-j} . It is easy to see that the f_j 's form a V -valued martingale on $[0, 1]$ with respect to Lebesgue measure and the σ -subalgebras of measurable sets generated by the partitions of $[0, 1]$ by dyadic intervals of length 2^{-j} , as in Section 103. Note that

$$(129.2) \quad \int_0^1 \|f_j(x)\| dx = \sum_{l=0}^{2^j-1} \|F((l+1)2^{-j}) - F(l2^{-j})\|.$$

If $F : [0, 1] \rightarrow V$ has finite length Λ , then

$$(129.3) \quad \int_0^1 \|f_j(x)\| dx \leq \Lambda$$

for each j . If F is Lipschitz, then the f_j 's are uniformly bounded. If F is differentiable at x , then

$$(129.4) \quad \lim_{j \rightarrow \infty} f_j(x) = F'(x).$$

If $V = L^1([0, 1])$ and $F(x)$ is the indicator function of $[0, x]$, then F is a Lipschitz function on $[0, 1]$ with values in $L^1([0, 1])$, as in Section 51. The corresponding martingale is the same as the one described in Section 100.

Now let $V = L^\infty(\mathbf{R})$, and let ϕ be a real or complex-valued Lipschitz function on the real line. Also let ϕ_x be the translate of ϕ by x , so that $\phi_x(y) = \phi(y-x)$. If ϕ is bounded, then $F(x) = \phi_x$ defines a Lipschitz mapping from \mathbf{R} into $L^\infty(\mathbf{R})$, as in Section 51. Otherwise, $F(x) = \phi_x - \phi$ defines a Lipschitz mapping from \mathbf{R} into $L^\infty(\mathbf{R})$, using only the hypothesis that ϕ is Lipschitz on \mathbf{R} . The restriction of $F(x)$ to $x \in [0, 1]$ defines a martingale $\{f_j\}_j$ with values in $L^\infty(\mathbf{R})$ as before. If ϕ is continuously-differentiable with uniformly continuous derivative, then F is differentiable at every $x \in \mathbf{R}$ as an $L^\infty(\mathbf{R})$ -valued function on \mathbf{R} . In this case, the derivative of F at x corresponds to -1 times the derivative of ϕ translated by x . If $x \in [0, 1]$, then it is easy to see that $\{f_j(x)\}_{j=1}^\infty$ converges to the same limit in $L^\infty(\mathbf{R})$. Conversely, if $\{f_j(x)\}_{j=1}^\infty$ converges in $L^\infty(\mathbf{R})$ for any $x \in [0, 1]$, then one can show that ϕ is continuously differentiable with uniformly continuous derivative. This is analogous to the fact that ϕ is continuously-differentiable with uniformly continuous derivative when F is differentiable at a single point, but slightly more complicated, since we are only using "dyadic" difference quotients of F . If $\{f_j(x)\}_{j=1}^\infty$ converges in $L^\infty(\mathbf{R})$ for some $x \in [0, 1]$, then the limit determines a bounded uniformly continuous function ψ on \mathbf{R} , because $f_j(x)$ corresponds to a bounded Lipschitz function on \mathbf{R} for each j that converges uniformly on \mathbf{R} as $j \rightarrow \infty$. One can check that $\psi = -\phi'_x$ where ϕ_x is differentiable, and then use the fact that Lipschitz functions are differentiable almost everywhere and can be represented by integrals of their derivatives to show that ϕ_x is continuously differentiable with derivative $-\psi$. Alternatively, one can argue that $\phi'_x = -\psi$ in the sense of distributions, and hence that ϕ_x is continuously differentiable with derivative $-\psi$.

Of course, one can just as well take V to be the space $C_b(\mathbf{R})$ of bounded continuous functions on the real line with the supremum norm here, which can be identified with a closed linear subspace of $L^\infty(\mathbf{R})$. There is also a simple way to embed $C_b(\mathbf{R})$ linearly and isometrically into ℓ^∞ , by restricting a bounded continuous function on the real line to the rationals, and then enumerating the latter by a sequence to get bounded sequences of real or complex numbers. If ϕ has compact support, then one can view to restriction of $F(x)$ to $x \in [0, 1]$ as a Lipschitz mapping into the space of continuous functions on a sufficiently large closed interval in the real line.

130 L^∞ Norms

Let (Y, \mathcal{B}, ν) be a probability space, and suppose that $g \in L^\infty(Y)$, so that

$$(130.1) \quad \|g\|_p = \left(\int_Y |g(y)|^p d\nu(y) \right)^{1/p} \leq \|g\|_\infty$$

for each $p < \infty$. Of course, $\|g\|_p$ is monotone increasing in p , by Jensen's inequality, and it is well known and not difficult to show that

$$(130.2) \quad \lim_{p \rightarrow \infty} \|g\|_p = \|g\|_\infty.$$

Similarly, if g is a measurable function on Y that is not essentially bounded, and if $g \in L^p(Y)$ for each $p < \infty$, then $\|g\|_p \rightarrow \infty$ as $p \rightarrow \infty$.

Let (X, \mathcal{A}, ν) be another probability space, and consider their Cartesian product $X \times Y$, equipped with the product measure $\mu \times \nu$. If $F(x, y)$ is a measurable function on $X \times Y$, then

$$(130.3) \quad N_\infty(F)(x) = \lim_{p \rightarrow \infty} N_p(F)(x) = \lim_{p \rightarrow \infty} \left(\int_Y |F(x, y)|^p d\nu(y) \right)^{1/p}$$

is a convenient way to express the norm of $F(x, y)$ as a function of y in $L^\infty(Y)$ for each $x \in X$. More precisely, it is often helpful to restrict p to be a positive integer here, so that $N_\infty(f)(x)$ is expressed as the limit of a monotone increasing sequence of functions. This makes it easy to derive properties of $N_\infty(F)(x)$ like those for $N_p(F)(x)$ when $p < \infty$ discussed earlier.

If $f(x) = \{f_l(x)\}_{l=1}^\infty$ is a sequence of real or complex-valued measurable functions on X , then the ℓ^∞ norm of $f(x)$ can be expressed as

$$(130.4) \quad \|f(x)\|_\infty = \sup_{l \geq 1} |f_l(x)| = \lim_{n \rightarrow \infty} \max_{1 \leq l \leq n} |f_l(x)|,$$

which implies that $\|f(x)\|_\infty$ is measurable on X . If $\|f(x)\|_\infty$ is integrable on X , then it is very easy to see that the ℓ^∞ norm of the conditional expectation of the $f_l(x)$'s with respect to some σ -subalgebra of \mathcal{A} is bounded by the conditional expectation of $\|f(x)\|_\infty$. One can simply use the fact that $|f_l(x)| \leq \|f(x)\|_\infty$ for each l to get that the conditional expectation of $f_l(x)$ is bounded by the conditional expectation of $\|f(x)\|_\infty$ for each l , and then take the supremum over l .

131 Paths and measures

Let $(V, \|v\|)$ be a real or complex Banach space, and let $F : [a, b] \rightarrow V$ be a path of finite length. As in Section 45, the one-sided limit $F(x+) = \lim_{y \rightarrow x+} F(y)$ exists for every $x \in [a, b)$, and similarly $F(x-) = \lim_{y \rightarrow x-} F(y)$ exists for every $x \in (a, b]$. We can extend F to the whole real line by putting $F(x) = F(a)$ when $x < a$ and $F(x) = F(b)$ when $x > b$, so that $F(a-) = F(a)$ and $F(b+) = F(b)$.

As in Section 44, we can put

$$(131.1) \quad \nu((r, t)) = F(t-) - F(r+)$$

when $a \leq r < t \leq b$, and

$$(131.2) \quad \nu([r, t]) = F(t+) - F(r-)$$

when $a \leq r \leq t \leq b$. Similarly, we can put

$$(131.3) \quad \nu([r, t)) = F(t-) - F(t+), \quad \nu((r, t]) = F(t+) - F(r-)$$

when $a \leq r < t \leq b$. This determines a finitely-additive V -valued measure on the algebra \mathcal{E} of subsets of $[a, b]$ that can be expressed as the union of finitely many intervals, where the intervals may be open, closed, or half-open and half-closed. Of course, this is a bit simpler when F is continuous.

Let $\alpha(x)$ be the length of the restriction of F to $[a, x]$ when $a \leq x \leq b$. This can be extended to all $x \in \mathbf{R}$ by setting $\alpha(x) = 0$ when $x < a$ and $\alpha(x) = \alpha(b)$ when $x > b$. Thus $\alpha(x)$ is a monotone increasing function on \mathbf{R} , which determines a nonnegative Borel measure μ on \mathbf{R} as in Section 44. It is easy to see that

$$(131.4) \quad \|\nu(A)\| \leq \mu(A)$$

for every $A \in \mathcal{E}$, because

$$(131.5) \quad \|F(t) - F(r)\| \leq \alpha(t) - \alpha(r)$$

when $r \leq t$. Note that $\alpha(t) - \alpha(r)$ is the same as the length of the restriction of F to $[r, t]$ when $r \leq t$, as in Section 41.

Let \mathcal{B} be the σ -algebra of Borel subsets of $[a, b]$. Thus $\mathcal{E} \subseteq \mathcal{B}$, and \mathcal{B} is the smallest σ -algebra of subsets of $[a, b]$ that contains \mathcal{E} . If $d(A, B) = \mu(A \triangle B)$ is the distance between $A, B \in \mathcal{B}$ associated to μ as in Section 79, then it follows that the closure of \mathcal{E} in \mathcal{B} with respect to $d(A, B)$ is equal to \mathcal{B} . This can also be seen more directly from the construction of μ .

If $A, B \in \mathcal{E}$, then

$$(131.6) \quad \begin{aligned} \nu(A) - \nu(B) &= (\nu(A \setminus B) + \nu(A \cap B)) - (\nu(B \setminus A) + \nu(A \cap B)) \\ &= \nu(A \setminus B) - \nu(B \setminus A), \end{aligned}$$

and hence

$$(131.7) \quad \begin{aligned} \|\nu(A) - \nu(B)\| &\leq \|\nu(A \setminus B)\| + \|\nu(B \setminus A)\| \\ &\leq \mu(A \setminus B) + \mu(B \setminus A) = d(A, B), \end{aligned}$$

by (131.4). This permits ν to be extended to a V -valued function on \mathcal{B} , using uniform continuity and completeness. More precisely, if $A \in \mathcal{B}$, then there is a sequence $\{A_j\}_{j=1}^{\infty}$ of elements of \mathcal{E} that converges to A with respect to $d(A, B)$. This implies that $\{\nu(A_j)\}_{j=1}^{\infty}$ is a Cauchy sequence in V , because of

the uniform continuity of ν with respect to $d(\cdot, \cdot)$ just established. It follows that $\{\nu(A_j)\}_{j=1}^\infty$ converges in V , because V is complete, and $\nu(A)$ is defined to be the limit of this sequence. One can also check that this does not depend on the particular sequence $\{A_j\}_{j=1}^\infty$ converging to A , using the uniform continuity of ν with respect to $d(\cdot, \cdot)$ again. Note that this extension satisfies

$$(131.8) \quad \|\nu(A) - \nu(B)\| \leq d(A, B)$$

for every $A, B \in \mathcal{B}$, since this holds when $A, B \in \mathcal{E}$ and is preserved under limits. In particular, (131.4) holds for every $A \in \mathcal{B}$.

Let $A, B \in \mathcal{B}$ be given, and let $\{A_j\}_{j=1}^\infty, \{B_j\}_{j=1}^\infty$ be sequences of elements of \mathcal{B} that converge to A, B with respect to $d(\cdot, \cdot)$, respectively. This implies that $\{A_j \cap B_j\}_{j=1}^\infty$ converges to $A \cap B$, and that $\{A_j \cup B_j\}_{j=1}^\infty$ converges to $A \cup B$, as in Section 79. Of course,

$$(131.9) \quad \nu(A_j) + \nu(B_j) = \nu(A_j \cap B_j) + \nu(A_j \cup B_j)$$

for each j , because ν is finitely additive on \mathcal{E} . Taking the limit as $j \rightarrow \infty$, we get that

$$(131.10) \quad \nu(A) + \nu(B) = \nu(A \cap B) + \nu(A \cup B),$$

because of (131.8). This shows that ν is finitely additive on \mathcal{B} .

If E_1, E_2, \dots is a sequence of elements of \mathcal{B} that are pairwise-disjoint, then

$$(131.11) \quad \sum_{l=1}^{\infty} \|\nu(E_l)\| \leq \sum_{l=1}^{\infty} \mu(E_l) = \mu\left(\bigcup_{l=1}^{\infty} E_l\right),$$

since (131.4) holds for every $A \in \mathcal{B}$. Moreover,

$$(131.12) \quad \left\| \nu\left(\bigcup_{l=n+1}^{\infty} E_l\right) \right\| \leq \mu\left(\bigcup_{l=n+1}^{\infty} E_l\right) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

This implies that

$$(131.13) \quad \sum_{l=1}^{\infty} \nu(E_l) = \nu\left(\bigcup_{l=1}^{\infty} E_l\right),$$

because we already know that ν is finitely additive on \mathcal{B} .

132 Paths and integrals

Let $(V, \|\cdot\|)$ be a real or complex Banach space, and let $F : [a, b] \rightarrow V$ be a path of finite length, as in the previous section. Also let ϕ be a continuous real or complex-valued function on $[a, b]$, as appropriate. Suppose that $\mathcal{P} = \{t_j\}_{j=0}^n$ is a partition of $[a, b]$, and that $t_{j-1} \leq r_j \leq t_j$ for $j = 1, \dots, n$, and consider

$$(132.1) \quad \sum_{j=1}^n \phi(r_j) (F(t_j) - F(t_{j-1})).$$

This is an approximation to the Riemann–Stieltjes integral of ϕ with respect to F , whose existence and basic properties will be discussed now. Basically, this is very similar to the Riemann–Stieltjes integral of a continuous function with respect to a real or complex-valued function of bounded variation on $[a, b]$.

If $t_{j-1} \leq r'_j \leq t_j$ is another collection of intermediate points, then the difference of the corresponding sums can be expressed as

$$(132.2) \quad \sum_{j=1}^n \phi(r_j) (F(t_j) - F(t_{j-1})) - \sum_{j=1}^n \phi(r'_j) (F(t_j) - F(t_{j-1})) \\ = \sum_{j=1}^n (\phi(r_j) - \phi(r'_j)) (F(t_j) - F(t_{j-1})).$$

Of course, ϕ is uniformly continuous on $[a, b]$, since it is continuous and $[a, b]$ is compact. Thus for each $\epsilon > 0$ there is a $\delta > 0$ such that

$$(132.3) \quad |\phi(r) - \phi(r')| \leq \epsilon$$

when $r, r' \in [a, b]$ and $|r - r'| < \delta$. In particular,

$$(132.4) \quad \left\| \sum_{j=1}^n \phi(r_j) (F(t_j) - F(t_{j-1})) - \sum_{j=1}^n \phi(r'_j) (F(t_j) - F(t_{j-1})) \right\| \\ \leq \sum_{j=1}^n |\phi(r_j) - \phi(r'_j)| \|F(t_j) - F(t_{j-1})\| \leq \epsilon \Lambda_a^b$$

when the mesh size $\max_{1 \leq j \leq n} (t_j - t_{j-1})$ of \mathcal{P} is strictly less than δ , where Λ_a^b denotes the length of F on $[a, b]$.

If $\mathcal{P}, \tilde{\mathcal{P}}$ are two partitions of $[a, b]$ with sufficiently small mesh size, then one can check that the difference between the corresponding sums (132.1) is also small. As usual, it is helpful to let $\hat{\mathcal{P}}$ be a common refinement of \mathcal{P} and $\tilde{\mathcal{P}}$, and to look at the differences between the sums corresponding to $\mathcal{P}, \tilde{\mathcal{P}}$ and the sum corresponding to $\hat{\mathcal{P}}$. These differences can be estimated in much the same way as in the previous paragraph, using the uniform continuity of ϕ . If $\mathcal{P}_1, \mathcal{P}_2, \dots$ is a sequence of partitions of $[a, b]$ whose mesh sizes are converging to 0, then the corresponding sums form a Cauchy sequence in V , and hence converges, by completeness of V . The limit does not depend on the particular sequence of partitions, because the difference between the sums associated to partitions with small mesh size is small, as before.

The Riemann–Stieltjes integral

$$(132.5) \quad \int_a^b \phi dF$$

of ϕ with respect to F is the limit of the sums (132.1) described in the previous paragraph. Observe that

$$(132.6) \quad \left\| \sum_{j=1}^n \phi(r_j) (F(t_j) - F(t_{j-1})) \right\| \leq \left(\sup_{a \leq r \leq b} |\phi(r)| \right) \Lambda_a^b$$

for every partition \mathcal{P} of $[a, b]$, and hence

$$(132.7) \quad \left\| \int_a^b \phi dF \right\| \leq \left(\sup_{a \leq r \leq b} |\phi(r)| \right) \Lambda_a^b.$$

If $\alpha(x)$ is the length of the restriction of F to $[a, x]$ for each $x \in [a, b]$, then one can improve this to get that

$$(132.8) \quad \left\| \int_a^b \phi dF \right\| \leq \int_a^b |\phi| d\alpha,$$

where the right side is a classical Riemann-Stieltjes integral. This is a more localized version of (132.7), which can be derived using the analogue of (132.7) on small subintervals of $[a, b]$. As in Section 44, the Riemann-Stieltjes integral of a continuous function on $[a, b]$ with respect to α can be extended to the Lebesgue-Stieltjes integral with respect to a positive Borel measure μ_α on $[a, b]$. As usual, continuous functions on $[a, b]$ form a dense linear subspace of $L^1(\mu_\alpha)$. Using (132.8), the Riemann-Stieltjes integral of ϕ with respect to F can be extended to $\phi \in L^1(\mu_\alpha)$. More precisely, if ϕ is an integrable function on $[a, b]$ with respect to μ_α , then there is a sequence $\{\phi_j\}_{j=1}^\infty$ of continuous functions on $[a, b]$ which converge to ϕ in $L^1(\mu_\alpha)$. Because of (132.8), the corresponding sequence of Riemann-Stieltjes integrals of the ϕ_j 's with respect to F form a Cauchy sequence in V , and therefore converges, by completeness. One can also check that the limit depends only on ϕ , and not on the particular sequence of continuous approximations $\{\phi_j\}_{j=1}^\infty$. Hence the Lebesgue-Stieltjes integral of ϕ with respect to F may be defined as this limit in V . Of course, this is very similar to the argument in the previous section.

133 Integrating vector measures

Let (X, \mathcal{A}) be a measurable space, and let $(V, \|v\|)$ be a real or complex Banach space. Also let μ be a V -valued function on \mathcal{A} such that for any sequence A_1, A_2, \dots of pairwise-disjoint measurable subsets of X ,

$$(133.1) \quad \sum_{j=1}^{\infty} \|\mu(A_j)\|$$

converges, and

$$(133.2) \quad \sum_{j=1}^{\infty} \mu(A_j) = \mu\left(\bigcup_{j=1}^{\infty} A_j\right).$$

As in Section 37, there is a nonnegative real-valued measure $\|\mu\|$ on X associated to μ such that

$$(133.3) \quad \|\mu(A)\| \leq \|\mu\|(A)$$

for each $A \in \mathcal{A}$, and $\|\mu\|(X) < \infty$.

Suppose that $f(x)$ is a real or complex-valued measurable simple function on X , as appropriate. This means that there are finitely many pairwise-disjoint measurable subsets A_1, \dots, A_n of X and real or complex numbers $\alpha_1, \dots, \alpha_n$ such that

$$(133.4) \quad f(x) = \sum_{j=1}^n \alpha_j \mathbf{1}_{A_j}(x).$$

Here $\mathbf{1}_A(x)$ is the indicator function associated to $A \subseteq X$ on X , equal to 1 when $x \in A$ and to 0 when $x \in X \setminus A$. The integral of f with respect to μ is given by

$$(133.5) \quad \int_X f \, d\mu = \sum_{j=1}^n \alpha_j \mu(A_j),$$

and satisfies

$$(133.6) \quad \left\| \int_X f \, d\mu \right\| \leq \sum_{j=1}^n |\alpha_j| \|\mu(A_j)\| = \int_X |f| \, d\|\mu\|.$$

More precisely, (133.5) does not depend on the particular representation (133.4) of f , and it also works when the A_j 's are not pairwise disjoint.

Let $f(x)$ be an integrable real or complex-valued function on X with respect to $\|\mu\|$, as appropriate, and let $\{f_l\}_{l=1}^\infty$ be a sequence of measurable simple functions on X that converge to f in $L^1(X, \|\mu\|)$. Using (133.6), one can check that

$$(133.7) \quad \left\{ \int_X f_l \, d\mu \right\}_{l=1}^\infty$$

is a Cauchy sequence in V , and hence converges, by completeness. The integral of f with respect to μ can be defined by

$$(133.8) \quad \int_X f \, d\mu = \lim_{l \rightarrow \infty} \int_X f_l \, d\mu.$$

As usual, one can also check that this does not depend on the sequence $\{f_l\}_{l=1}^\infty$ of simple functions converging to f , and that

$$(133.9) \quad \left\| \int_X f \, d\mu \right\| \leq \int_X |f| \, d\|\mu\|,$$

by (133.6).

If λ is a bounded linear functional on V , then

$$(133.10) \quad \mu_\lambda(A) = \lambda(\mu(A))$$

defines a real or complex measure on X , as appropriate. Note that

$$(133.11) \quad |\mu_\lambda(A)| = |\lambda(\mu(A))| \leq \|\lambda\|_* \|\mu(A)\| \leq \|\lambda\|_* \|\mu\|(A),$$

and hence

$$(133.12) \quad |\mu_\lambda|(A) \leq \|\lambda\|_* \|\mu\|(A)$$

for every $A \in \mathcal{A}$. If f is a measurable simple function on X , then it is easy to see that

$$(133.13) \quad \lambda\left(\int_X f d\mu\right) = \int_X f d\mu_\lambda$$

for every $\lambda \in V^*$. This also works when $f \in L^1(X, \|\mu\|)$, by approximating f by simple functions, as in the previous paragraph. The integral of f with respect to μ is uniquely determined by this property, because of the Hahn–Banach theorem.

134 Measures and orthogonality

Let (X, \mathcal{A}) be a measurable space, and let $(V, \langle v, w \rangle)$ be a real or complex Hilbert space. Also let $\nu(A)$ be a finitely-additive V -valued measure on (X, \mathcal{A}) such that

$$(134.1) \quad \langle \nu(A), \nu(B) \rangle = 0$$

whenever A, B are disjoint measurable subsets of X . In particular,

$$(134.2) \quad \|\nu(A \cup B)\|^2 = \|\nu(A)\|^2 + \|\nu(B)\|^2$$

when A, B are disjoint. It follows that

$$(134.3) \quad \sum_{j=1}^n \|\nu(A_j)\|^2 + \left\| \nu\left(\bigcup_{j=n+1}^{\infty} A_j\right) \right\|^2 = \left\| \nu\left(\bigcup_{j=1}^{\infty} A_j\right) \right\|^2$$

for any sequence A_1, A_2, \dots of pairwise-disjoint measurable subsets of X and $n \geq 1$, and hence

$$(134.4) \quad \sum_{j=1}^n \|\nu(A_j)\|^2 \leq \left\| \nu\left(\bigcup_{j=1}^{\infty} A_j\right) \right\|^2.$$

Thus

$$(134.5) \quad \sum_{j=1}^{\infty} \|\nu(A_j)\|^2 \leq \left\| \nu\left(\bigcup_{j=1}^{\infty} A_j\right) \right\|^2,$$

which implies that $\sum_{j=1}^{\infty} \nu(A_j)$ converges in V when A_1, A_2, \dots are disjoint. In this case, we ask also that

$$(134.6) \quad \sum_{j=1}^{\infty} \nu(A_j) = \nu\left(\bigcup_{j=1}^{\infty} A_j\right),$$

which implies that

$$(134.7) \quad \sum_{j=1}^{\infty} \|\nu(A_j)\|^2 = \left\| \nu\left(\bigcup_{j=1}^{\infty} A_j\right) \right\|^2.$$

This shows that $\|\nu(A)\|^2$ is a nonnegative real-valued measure on X under these conditions, which may be denoted $\|\nu\|^2$.

As a basic example of this type of situation, let μ be a nonnegative real-valued measure on X , and consider $V = L^2(X, \mu)$, with the standard integral inner product. Let $g \in L^2(X, \mu)$ be given, and let ν_g be the $L^2(X, \mu)$ -valued function on \mathcal{A} defined by

$$(134.8) \quad \nu_g(A) = g \mathbf{1}_A.$$

Equivalently, $\nu_g(A)$ is the function on X equal to g on A and to 0 on $X \setminus A$ for each measurable set $A \subseteq X$. In particular,

$$(134.9) \quad \|\nu_g(A)\|^2 = \int_A |g|^2 d\mu.$$

It is easy to see that ν_g satisfies all of the conditions described in the previous paragraph.

Let V be any Hilbert space again, and let ν be a V -valued function on \mathcal{A} that satisfies the same conditions as before. Let A_1, \dots, A_n be finitely many pairwise-disjoint measurable subsets of X , and let $\alpha_1, \dots, \alpha_n$ be real or complex numbers, as appropriate. If $f = \sum_{j=1}^n \alpha_j \mathbf{1}_{A_j}$ is the corresponding simple function, then its integral with respect to ν is given by

$$(134.10) \quad \int_X f d\nu = \sum_{j=1}^n \alpha_j \nu(A_j).$$

In this case,

$$(134.11) \quad \left\| \int_X f d\nu \right\|^2 = \sum_{j=1}^n |\alpha_j|^2 \|\nu(A_j)\|^2 = \int_X |f|^2 d\|\nu\|^2.$$

Using standard arguments based on continuity and completeness, the integral of f with respect to ν can be extended to an isometric linear mapping from $L^2(X, \|\nu\|^2)$ into V .

Suppose that $V = L^2(X, \mu)$ for some nonnegative real-valued measure μ on X , and that $\nu = \nu_g$ for some $g \in L^2(X, \mu)$. If f is a measurable simple function on X , then it is easy to see that

$$(134.12) \quad \int_X f d\nu_g = f g$$

as an element of $L^2(X, \mu)$, and that

$$(134.13) \quad \int_X |f|^2 d\|\nu_g\|^2 = \int_X |f|^2 |g|^2 d\mu.$$

If $f \in L^2(X, \|\nu_g\|^2)$, then $f g \in L^2(X, \mu)$, and the same statements hold.

135 Paths and orthogonality

Let $(V, \langle v, w \rangle)$ be a real or complex Hilbert space, and let $p(t)$ be a V -valued function on a closed interval $[a, b]$ in the real line. Suppose that

$$(135.1) \quad \langle p(t_2) - p(t_1), p(t_3) - p(t_2) \rangle = 0$$

whenever $a \leq t_1 \leq t_2 \leq t_3 \leq b$, which implies that

$$(135.2) \quad \langle p(t_2) - p(t_1), p(t_4) - p(t_3) \rangle = 0$$

when $t_3 \leq t_4 \leq b$ too. More precisely, (135.1) also holds with t_3 replaced by t_4 in this case, and (135.2) follows by expressing $p(t_4) - p(t_3)$ as the difference of $p(t_4) - p(t_2)$ and $p(t_3) - p(t_2)$. If we put

$$(135.3) \quad \alpha(t) = \|p(t) - p(a)\|^2$$

for $a \leq t \leq b$, then

$$(135.4) \quad \begin{aligned} \alpha(t) &= \|p(r) - p(a)\|^2 + \|p(t) - p(r)\|^2 \\ &= \alpha(r) + \|p(t) - p(r)\|^2 \geq \alpha(r) \end{aligned}$$

when $a \leq r \leq t \leq b$, so that $\alpha(t)$ is monotone increasing on $[a, b]$. One can show that the one-sided limit $p(t+)$ exists when $a \leq t < b$, and similarly that $p(t-)$ exists when $a < t \leq b$, in analogy with Section 45. Note that $p(t)$ is continuous at the same points where $\alpha(t)$ is continuous, because of (135.4). It is convenient to extend $p(t)$ to the whole real line, by putting $p(t) = p(a)$ when $t < a$ and $p(t) = p(b)$ when $t > b$, so that $p(a-) = p(a)$ and $p(b+) = p(b)$ are defined as well. We can extend $\alpha(t)$ to \mathbf{R} in the same way, so that $\alpha(t) = 0$ when $t < a$ and $\alpha(t) = \alpha(b)$ when $t > b$.

As in Sections 44 and 131, put

$$(135.5) \quad \nu((r, t)) = p(t-) - p(r+)$$

and

$$(135.6) \quad \nu([r, t)) = p(t-) - p(r-), \quad \nu((r, t]) = p(t+) - p(r+)$$

when $a \leq r < t \leq b$, and

$$(135.7) \quad \nu([r, t]) = p(t+) - p(r-)$$

when $a \leq r \leq t \leq b$. This determines a finitely-additive V -valued measure on the algebra \mathcal{E} of subsets of $[a, b]$ that can be expressed as the union of finitely many intervals, where the intervals may be open, closed, or half-open and half-closed. By hypothesis,

$$(135.8) \quad \langle \nu(I), \nu(I') \rangle = 0$$

for every pair I, I' of disjoint subintervals of $[a, b]$. If μ_α is the nonnegative Borel measure associated to $\alpha(t)$ as in Section 44, then

$$(135.9) \quad \|\nu(A)\|^2 = \mu_\alpha(A)$$

for every subinterval A of $[a, b]$. This also works when $A \in \mathcal{E}$, because A is then the union of finitely many pairwise-disjoint subintervals I_1, \dots, I_n of $[a, b]$, and $\nu(I_1), \dots, \nu(I_n)$ are orthogonal to each other in V .

Let

$$(135.10) \quad f(t) = \sum_{j=1}^n c_j \mathbf{1}_{I_j}(t)$$

be a step function on $[a, b]$, where I_1, \dots, I_n are pairwise-disjoint subintervals of $[a, b]$, and c_1, \dots, c_n are real or complex numbers, as appropriate. The integral of f with respect to ν can be defined by

$$(135.11) \quad \int_a^b f \, d\nu = \sum_{j=1}^n c_j \nu(I_j).$$

In this case,

$$(135.12) \quad \left\| \int_a^b f \, d\nu \right\|^2 = \sum_{j=1}^n |c_j|^2 \|\nu(I_j)\|^2,$$

because $\nu(I_1), \dots, \nu(I_n)$ are orthogonal to each other in V . Hence

$$(135.13) \quad \left\| \int_a^b f \, d\nu \right\|^2 = \int_a^b |f|^2 \, d\mu_\alpha,$$

as in (135.9). Thus the integral of f with respect to ν defines a linear isometry from the subspace of $L^2([a, b], \mu_\alpha)$ consisting of step functions into V . This can be extended to a linear isometry from $L^2([a, b], \mu_\alpha)$ into V , by standard arguments of continuity and completeness. In particular, ν can be extended to a V -valued Borel measure on $[a, b]$ as in the previous section, by applying this extension to indicator functions of measurable subsets of $[a, b]$.

If μ is a finite nonnegative Borel measure on $[a, b]$, then $p(t) = \mathbf{1}_{[a, t]}$ defines a mapping from $[a, b]$ into $L^2([a, b], \mu)$ that satisfies the conditions mentioned at the beginning of the section. One could also use the indicator function associated to (a, t) , $[a, t)$, or $(a, t]$, and the corresponding differences of one-sided limits of p would be the same. Note that these indicator functions are already the same in $L^2([a, b], \mu)$ when $\mu(\{x\}) = 0$ for each $x \in [a, b]$, in which case p is continuous. One can check that $\mu_\alpha = \mu$ in this situation, and that the embedding described in the preceding paragraph reduces to the identity mapping on $L^2([a, b], \mu)$.

136 Minkowski's integral inequality

Let (X, \mathcal{A}, μ) , (Y, \mathcal{B}, ν) be measure spaces, with finite or σ -finite measure. If $F(x, y)$ is a nonnegative measurable function on the Cartesian product $X \times Y$ and $1 \leq p < \infty$, then *Minkowski's integral inequality* states that

$$(136.1) \quad \begin{aligned} & \left(\int_Y \left(\int_X F(x, y) \, d\mu(x) \right)^p \, d\nu(y) \right)^{1/p} \\ & \leq \int_X \left(\int_Y F(x, y)^p \, d\nu(y) \right)^{1/p} \, d\mu(x). \end{aligned}$$

This is an integrated version of the triangle inequality for the L^p norm, which is also known as Minkowski's inequality. Note that one has equality in (136.1) when $p = 1$, by Fubini's theorem. We have basically encountered versions of this already in connection with conditional expectation, and we would like to mention a couple of other approaches now.

Let A_1, \dots, A_n be finitely many pairwise-disjoint measurable subsets of X whose union is equal to X . If $F(x, y)$ is constant in x on each A_j , then (136.1) reduces to the ordinary Minkowski inequality for finite sums. Otherwise, one can get (136.1) by approximating $F(x, y)$ by functions of this type. This is analogous to the earlier discussion of "nice functions" on $X \times Y$, but with the roles of X and Y exchanged. A key point is that measurable subsets of $X \times Y$ with finite measure can be approximated by finite unions of measurable rectangles, as in Section 123.

Alternatively, put

$$(136.2) \quad N_p(F)(x) = \left(\int_Y F(x, y)^p d\nu(y) \right)^{1/p},$$

as before. If μ is a probability measure on X , then

$$(136.3) \quad \left(\int_X F(x, y) d\mu(x) \right)^p \leq \int_X F(x, y)^p d\mu(x)$$

for each $y \in Y$, by Jensen's inequality. Hence

$$(136.4) \quad \begin{aligned} \int_Y \left(\int_X F(x, y) d\mu(x) \right)^p d\nu(y) &\leq \int_Y \int_X F(x, y)^p d\mu(x) d\nu(y) \\ &= \int_X N_p(F)(x)^p d\mu(x), \end{aligned}$$

by Fubini's theorem. If $N_p(F)(x) \leq 1$ for μ -almost every $x \in X$, then it follows that

$$(136.5) \quad \int_Y \left(\int_X F(x, y) d\mu(x) \right)^{1/p} d\nu(y) \leq 1.$$

This may be considered as a special case of (136.1), and the general case may be derived from it using homogeneity, as follows. If the right side of (136.1) is equal to 0, then $F(x, y) = 0$ almost everywhere on $X \times Y$, the left side of (136.1) is also equal to 0, and there is nothing to do. There is also nothing to do when the right side of (136.1) is $+\infty$. Thus we may suppose that the right side of (136.1) is positive and finite, and we can even take it to be equal to 1, by multiplying F by a positive constant. We may also suppose that $N_p(F)(x) > 0$ for every $x \in X$, because the $x \in X$ for which $N_p(F)(x) = 0$ do not play a role in (136.1). If we put

$$(136.6) \quad F'(x, y) = N_p(x)^{-1} F(x, y),$$

then $N_p(F')(x) = 1$ for every $x \in X$ automatically. Similarly, if we put

$$(136.7) \quad \mu'(A) = \int_A N_p(F)(x) d\mu(x),$$

then μ' is a probability measure on X , because the right side of (136.1) is supposed to be equal to 1. The special case of Minkowski's integral inequality under consideration implies that

$$(136.8) \quad \int_Y \left(\int_X F'(x, y) d\mu'(x) \right)^p d\nu(y) \leq 1.$$

This implies that the left side of (136.1) is less than or equal to 1, as desired.

Let $N_\infty(F)(x)$ be the essential supremum of $F(x, y)$ over $y \in Y$. The $p = \infty$ version of (136.1) states that the essential supremum of

$$(136.9) \quad \int_X F(x, y) d\mu(y)$$

over $y \in Y$ is less than or equal to

$$(136.10) \quad \int_X N_\infty(F)(x) d\mu(x).$$

If Y is a probability space, then this can be obtained from (136.1) by taking the limit as $p \rightarrow \infty$ with $p \in \mathbf{Z}_+$, as in Section 130. Otherwise, one can reduce to the case of probability spaces by approximating Y by subsets of finite measure, or using a positive weight on Y with integral 1. Alternatively, if $N_\infty(F)(x) \leq 1$ for almost every $x \in X$, then $F(x, y) \leq 1$ for almost every $(x, y) \in X \times Y$, by Fubini's theorem. If μ is a probability measure on X , then it follows that (136.9) is less than or equal to 1 for almost every $y \in Y$. As in the previous paragraph, this may be considered as a special case of the desired estimate, and the general case can be derived from it in the same way as before.

137 Spaces of measures

Let (X, \mathcal{A}) be a measurable space, and let $(V, \|v\|)$ be a real or complex Banach space. Consider the space $\mathcal{M}(X, V)$ of V -valued functions μ on \mathcal{A} such that

$$(137.1) \quad \sum_{j=1}^{\infty} \|\mu(A_j)\| < \infty$$

and

$$(137.2) \quad \sum_{j=1}^{\infty} \mu(A_j) = \mu\left(\bigcup_{j=1}^{\infty} A_j\right)$$

for every sequence A_1, A_2, \dots of pairwise-disjoint measurable subsets of X . As usual, the first condition already implies that $\sum_{j=1}^{\infty} \mu(A_j)$ converges in V . The second condition is equivalent to asking that μ be finitely additive and have the continuity property that

$$(137.3) \quad \lim_{n \rightarrow \infty} \mu\left(\bigcup_{j=1}^n A_j\right) = \mu\left(\bigcup_{j=1}^{\infty} A_j\right),$$

just as for real or complex measures.

Note that $\mathcal{M}(X, V)$ is a vector space over the real or complex numbers, as appropriate. If $\mu \in \mathcal{M}(X, V)$, then $p(A) = \|\mu(A)\|$ satisfies the conditions in Section 35, and $\|\mu\|(A) = p^*(A)$ is a finite nonnegative measure on X , as in Section 37. By construction,

$$(137.4) \quad \|\mu(A)\| \leq \|\mu\|(A)$$

for every measurable set $A \subseteq X$, and $\|\mu\|(A)$ is the smallest nonnegative measure on X with this property, as in Section 35. It is easy to check that $\|\mu\|(X)$ defines a norm on $\mathcal{M}(X, V)$.

Suppose that μ_1, μ_2, \dots is a sequence of elements of $\mathcal{M}(X, V)$ which is a Cauchy sequence with respect to this norm. Thus for each $\epsilon > 0$ there is an $L \geq 1$ such that

$$(137.5) \quad \|\mu_l - \mu_n\|(X) < \epsilon$$

for every $l, n \geq L$. Of course,

$$(137.6) \quad \|\mu_l(A) - \mu_n(A)\| \leq \|\mu_l - \mu_n\|(A) \leq \|\mu_l - \mu_n\|(X)$$

for every measurable set $A \subseteq X$ and $l, n \geq 1$, which implies that $\{\mu_l(A)\}_{l=1}^\infty$ is a Cauchy sequence in V for every $A \in \mathcal{A}$. Let $\mu(A)$ be the limit of this sequence in V , which converges by completeness. Note that $\{\mu_l(A)\}_{l=1}^\infty$ actually converges to $\mu(A)$ uniformly on \mathcal{A} , because the Cauchy condition holds uniformly over $A \in \mathcal{A}$.

If A_1, A_2, \dots is a sequence of pairwise-disjoint measurable subsets of X , then

$$(137.7) \quad \sum_{j=1}^\infty \|\mu_l(A_j)\| \leq \sum_{j=1}^\infty \|\mu_l\|(A_j) = \|\mu_l\|\left(\bigcup_{j=1}^\infty A_j\right) \leq \|\mu_l\|(X)$$

for each l . In the limit as $l \rightarrow \infty$, we get that

$$(137.8) \quad \sum_{j=1}^\infty \|\mu(A_j)\| \leq \sup_{l \geq 1} \|\mu_l\|(X).$$

The right side is finite because $\{\mu_l\}_{l=1}^\infty$ is a Cauchy sequence, and hence is bounded. It is easy to see that $\mu(A)$ is finitely additive, since $\mu_l(A)$ is finitely additive for each l . The continuity condition (137.3) can also be derived from the corresponding property of the μ_l 's, using the fact that $\{\mu_l(A)\}_{l=1}^\infty$ converges to $\mu(A)$ uniformly on \mathcal{A} . Similarly, if A_1, A_2, \dots is a sequence of pairwise-disjoint measurable subsets of X whose union is equal to X , then

$$(137.9) \quad \sum_{j=1}^\infty \|\mu_l(A_j) - \mu_n(A_j)\| \leq \sum_{j=1}^\infty \|\mu_l(A_j) - \mu_n(A_j)\| = \|\mu_l - \mu_n\|(X)$$

for each $l, n \geq 1$, as before. This implies that

$$(137.10) \quad \sum_{j=1}^\infty \|\mu(A_j) - \mu_n(A_j)\| \leq \sup_{l \geq n} \|\mu_l - \mu_n\|(X)$$

for each $n \geq 1$, by taking the limit as $l \rightarrow \infty$, as in (137.8). It follows that

$$(137.11) \quad \|\mu - \mu_n\|(X) \leq \sup_{l \geq n} \|\mu_l - \mu_n\|(X)$$

for each n , by taking the supremum over all such partitions $\{A_j\}_{j=1}^\infty$ of X . This shows that $\mu \in \mathcal{M}(X, V)$ and that $\{\mu_n\}_{n=1}^\infty$ converges to μ with respect to the norm $\|\mu\|(X)$, and hence that $\mathcal{M}(X, V)$ is complete.

138 Products and measures

Let (X, \mathcal{A}, μ) , (Y, \mathcal{B}, ν) be finite or σ -finite measure spaces, and let $F(x, y)$ be a measurable function on $X \times Y$. As usual, we put

$$(138.1) \quad N_p(F)(x) = \left(\int_Y |F(x, y)|^p d\nu(y) \right)^{1/p}$$

when $1 \leq p < \infty$, and we let $N_\infty(F)(x)$ be the essential supremum of $|F(x, y)|$ over $y \in Y$. Suppose that

$$(138.2) \quad \int_X N_p(F)(x) d\mu(x) < \infty$$

for some p , $1 \leq p \leq \infty$, and put

$$(138.3) \quad \phi(A)(y) = \int_A F(x, y) d\mu(x)$$

for each measurable set $A \subseteq X$. This defines $\phi(A)$ as a measurable function on Y which is in $L^p(Y)$ and satisfies

$$(138.4) \quad \|\phi(A)\|_{L^p(Y)} \leq \int_A N_p(F)(x) d\mu(x),$$

by Minkowski's integral inequality. If A_1, A_2, \dots is a sequence of pairwise-disjoint measurable subsets of X , then

$$(138.5) \quad \begin{aligned} \sum_{j=1}^\infty \|\phi(A_j)\|_{L^p(Y)} &\leq \sum_{j=1}^\infty \int_{A_j} N_p(F)(x) d\mu(x) \\ &= \int_{\bigcup_{j=1}^\infty A_j} N_p(F)(x) d\mu(x) < \infty. \end{aligned}$$

Thus $\sum_{j=1}^\infty \phi(A_j)$ converges in $L^p(Y)$, and it is easy to see that

$$(138.6) \quad \sum_{j=1}^\infty \phi(A_j) = \phi\left(\bigcup_{j=1}^\infty A_j\right).$$

Hence $\phi \in \mathcal{M}(X, L^p(Y))$. If $\|\phi\|(A)$ is as in the previous section, then

$$(138.7) \quad \|\phi\|(A) \leq \int_A N_p(F)(x) d\mu(x),$$

because of (138.4).

139 L^p -Valued measures

Let (X, \mathcal{A}) be a measurable space, and let (Y, \mathcal{B}, ν) be a σ -finite measure space. Suppose that $\mu \in \mathcal{M}(X, L^p(Y))$ for some p , $1 < p \leq \infty$. Thus $\|\mu\|$ is a finite nonnegative real measure on X , and we can consider the product measure $\|\mu\| \times \nu$ on $X \times Y$. We would like to represent μ by a function on $X \times Y$, as in the previous section.

Let A_1, \dots, A_n be finitely many pairwise-disjoint measurable subsets of X such that $\bigcup_{j=1}^n A_j = X$, and let $g_1(y), \dots, g_n(y)$ be elements of $L^q(Y)$, where $1 \leq q < \infty$ is the exponent conjugate to p , $1/p + 1/q = 1$. Put $G(x, y) = g_j(y)$ when $x \in A_j$, and

$$(139.1) \quad L(G) = \sum_{j=1}^n \int_Y \mu(A_j)(y) g_j(y) d\nu(y).$$

By Hölder's inequality,

$$(139.2) \quad \left| \int_Y \mu(A_j)(y) g_j(y) d\nu(y) \right| \leq \|\mu(A_j)\|_{L^p(Y)} \|g_j\|_{L^q(Y)} \leq \|\mu\|(A_j) \|g_j\|_{L^q(Y)}$$

for each j . This implies that

$$(139.3) \quad |L(G)| \leq \int_X N_q(G)(x) d\|\mu\|,$$

where $N_q(G)(x)$ denotes the $L^q(Y)$ norm of $G(x, y)$ as a function of y , as usual. In particular,

$$(139.4) \quad |L(G)| \leq \|\mu\|(X)^{1/p} \|G\|_{L^q(X \times Y, \|\mu\| \times \nu)}.$$

It is easy to see that (139.1) does not depend on the particular representation of $G(x, y)$ in the preceding paragraph, because μ is finitely additive. One can also check that the collection of these functions $G(x, y)$ forms a linear subspace of $L^q(X \times Y, \|\mu\| \times \nu)$, and that $L(G)$ defines a linear functional on this subspace. The main point is that any two partitions of X into finitely many measurable sets has a common refinement, and so any two functions of this type can be represented in this way using the same partition of X . This subspace is also dense in $L^q(X \times Y, \|\mu\| \times \nu)$, because $q < \infty$. We also know from (139.4) that $L(G)$ is a bounded linear functional on this subspace, with respect to the L^q norm, and hence has a unique extension to a bounded linear functional on $L^q(X \times Y, \|\mu\| \times \nu)$.

The Riesz representation theorem implies that there is a unique element $F(x, y)$ of $L^p(X \times Y, \|\mu\| \times \nu)$ such that

$$(139.5) \quad L(G) = \int_{X \times Y} F(x, y) G(x, y) d\|\mu\|(x) d\nu(y)$$

for every $G \in L^q(X \times Y, \|\mu\| \times \nu)$. If A is a measurable subset of X and $g(y) \in L^q(Y)$, then we can apply this to $G(x, y) = \mathbf{1}_A(x) g(y)$, to get that

$$(139.6) \quad \int_Y \mu(A)(y) g(y) d\nu(y) = \int_Y \left(\int_A F(x, y) d\mu(x) \right) g(y) d\nu(y).$$

It follows that

$$(139.7) \quad \mu(A)(y) = \int_A F(x, y) d\mu(x)$$

as elements of $L^p(Y)$ for every measurable set $A \subseteq X$, as in the previous section. Moreover,

$$(139.8) \quad \|F\|_{L^p(X \times Y, \|\mu\| \times \nu)} \leq \|\mu\|(X)^{1/p},$$

because of (139.4). If $p = \infty$, then this says that the L^∞ norm of $F(x, y)$ is less than or equal to 1 on $X \times Y$. Otherwise, if $p < \infty$, and if A is a measurable subset of X , then (139.3) implies that

$$(139.9) \quad |L(G)| \leq \|\mu\|(A)^{1/p} \|G\|_{L^q(A \times Y, \|\mu\| \times \nu)}$$

when $G(x, y) = 0$ for every $x \in X \setminus A$. Hence

$$(139.10) \quad \left(\int_A \int_Y |F(x, y)|^p d\mu(x) d\nu(y) \right)^{1/p} \leq \|\mu\|(A)^{1/p},$$

or equivalently,

$$(139.11) \quad \int_A N_p(F)(x)^p d\mu(x) \leq \|\mu\|(A).$$

This shows that $N_p(F)(x) \leq 1$ almost everywhere on X with respect to $\|\mu\|$.

140 ℓ^1 -Valued measures

Let (X, \mathcal{A}) be a measurable space, and let μ_1, μ_2, \dots be a sequence of real or complex-valued measures on X such that $\sum_{j=1}^{\infty} |\mu_j|(X) < \infty$. This implies that

$$(140.1) \quad \sum_{j=1}^{\infty} |\mu_j(A)| \leq \sum_{j=1}^{\infty} |\mu_j|(A) \leq \sum_{j=1}^{\infty} |\mu_j|(X) < \infty$$

for every measurable set $A \subseteq X$, which means that $\mu(A) = \{\mu_j(A)\}_{j=1}^{\infty} \in \ell^1$ for each $A \in \mathcal{A}$. Put $\rho(A) = \sum_{j=1}^{\infty} |\mu_j|(A)$, so that ρ is a finite nonnegative real measure on X by hypothesis, and

$$(140.2) \quad \|\mu(A)\|_1 = \sum_{j=1}^{\infty} |\mu_j(A)| \leq \rho(A)$$

for each $A \in \mathcal{A}$. Using this, one can check that $\mu \in \mathcal{M}(X, \ell^1)$, and that $\|\mu\|(A) \leq \rho(A)$ for each $A \in \mathcal{A}$.

This construction is actually equivalent to the one in Section 138, with $p = 1$ and $Y = \mathbf{Z}_+$, equipped with counting measure. This is because μ_j is absolutely continuous with respect to ρ for each j , and hence can be expressed in terms of an integrable function f_j with respect to ρ , as in the Radon–Nikodym theorem. The L^1 norm of f_j with respect to ρ is equal to $|\mu_j|(X)$ for each j , and is summable over j . Thus the sequence of f_j 's can be identified with an integrable function on $X \times \mathbf{Z}_+$, using ρ as the measure on X .

Conversely, suppose that $\mu \in \mathcal{M}(X, \ell^1)$. Thus $\mu(A) = \{\mu_j(A)\}_{j=1}^\infty$ for some real or complex-valued functions μ_j on \mathcal{A} , as appropriate. It is easy to see that μ_j is a real or complex measure on X for each j , because of the corresponding properties of μ . A key point now is that

$$(140.3) \quad \sum_{j=1}^{\infty} |\mu_j|(A) \leq \|\mu\|(A)$$

for every $A \in \mathcal{A}$. Of course, it suffices to show that

$$(140.4) \quad \sum_{j=1}^n |\mu_j|(A) \leq \|\mu\|(A)$$

for every $A \in \mathcal{A}$ and $n \geq 1$. Remember that $|\mu_j|(A) = p_j^*(A)$ is defined as in Section 35, using $p_j(A) = |\mu_j(A)|$. More precisely, $p_j^*(A)$ can be defined as the supremum of sums of p_j over partitions of A into finitely many measurable subsets. If we use the same partition of A for each j , then the desired estimate would follow from the definition of $\|\mu\|(A)$ as $p^*(A)$ with $p(A) = \|\mu(A)\|_{\ell^1}$. If instead we have different partitions of A for $j = 1, \dots, n$, then we can use a common refinement of them to reduce to the case of a single partition of A .

Suppose now that $\mu \in \mathcal{M}(X, \ell^p)$, $1 \leq p \leq \infty$. As in the preceding paragraph, $\mu(A) = \{\mu_j(A)\}_{j=1}^\infty$, where each μ_j is a real or complex measure on X . It is easy to see that μ_j is absolutely continuous with respect to $\|\mu\|$ for each j , and so can be expressed in terms of an integrable function with respect to $\|\mu\|$, by the Radon–Nikodym theorem. If $p = 1$, then the L^1 norms of these functions are summable, as before. If $p > 1$, then we are back in the situation of the previous section, with $Y = \mathbf{Z}_+$ equipped with counting measure.

141 Finite sums

Let (X, \mathcal{A}) be a measurable space, and let $(V, \|v\|)$ be a real or complex Banach space. Suppose that μ_1, \dots, μ_n are finitely many real or complex measures on X , as appropriate, and that v_1, \dots, v_n are vectors in V . It is easy to see that

$$(141.1) \quad \mu(A) = \sum_{j=1}^n \mu_j(A) v_j$$

defines an element of $\mathcal{M}(X, V)$. Of course,

$$(141.2) \quad \|\mu(A)\| \leq \sum_{j=1}^n |\mu_j(A)| \|v_j\| \leq \sum_{j=1}^n |\mu_j|(A) \|v_j\|$$

for each $A \in \mathcal{A}$, which implies that

$$(141.3) \quad \|\mu\|(A) \leq \sum_{j=1}^n |\mu_j|(A) \|v_j\|.$$

Let ρ be a finite nonnegative real measure on X such that μ_j is absolutely continuous with respect to ρ for each j . One can take

$$(141.4) \quad \rho = \sum_{j=1}^n |\mu_j|,$$

for instance. By the Radon–Nikodym theorem, there are integrable functions f_1, \dots, f_n on X with respect to ρ such that

$$(141.5) \quad \mu_j(A) = \int_A f_j d\rho$$

for each $A \in \mathcal{A}$ and $j = 1, \dots, n$. If $f(x) = \sum_{j=1}^n f_j(x) v_j$, then

$$(141.6) \quad \|\mu(A)\| = \left\| \sum_{j=1}^n v_j \int_A f_j d\rho \right\| \leq \int_A \|f\| d\rho$$

for each $A \in \mathcal{A}$, as in Section 120. This implies that

$$(141.7) \quad \|\mu\|(A) \leq \int_A \|f\| d\rho$$

for each $A \in \mathcal{A}$.

More precisely,

$$(141.8) \quad \|\mu\|(A) = \int_A \|f\| d\rho$$

for each $A \in \mathcal{A}$ under these conditions. To see this, remember that

$$(141.9) \quad \sum_{k=1}^l \|\mu(A_k)\| \leq \|\mu\|(A)$$

when A_1, \dots, A_l are pairwise-disjoint measurable sets whose union is A , by definition of $\|\mu\|(A)$. In order to show that

$$(141.10) \quad \int_A \|f\| d\rho \leq \|\mu\|(A),$$

one can choose measurable sets A_k on which the f_j 's are approximately constant.

Let us now start with a measure $\mu \in \mathcal{M}(X, V)$ that takes values in a finite-dimensional linear subspace of V . If v_1, \dots, v_n is a basis for this linear subspace, then there are unique real or complex measures μ_1, \dots, μ_n on X for which μ can be expressed as in (141.1). Because any two norms on a finite-dimensional real or complex vector space are equivalent,

$$(141.11) \quad \left\| \sum_{j=1}^n t_j v_j \right\| \geq c \max_{1 \leq j \leq n} |t_j|$$

for some $c > 0$ and every $t_1, \dots, t_n \in \mathbf{R}$ or \mathbf{C} , as appropriate. This implies that

$$(141.12) \quad c \max_{1 \leq j \leq n} |\mu_j(A)| \leq \|\mu(A)\| \leq \|\mu\|(A)$$

for each $A \in \mathcal{A}$, and hence that μ_j is absolutely continuous with respect to $\|\mu\|$ for each j . Thus we can take $\rho = \|\mu\|$ in the previous paragraphs, and it follows that the corresponding function f satisfies $\|f(x)\| = 1$ for almost every $x \in X$ with respect to $\|\mu\|$.

142 Approximations

Let (X, \mathcal{A}) be a measurable space, and let $(V, \|v\|)$ be a real or complex Banach space. Suppose that μ_1, μ_2, \dots is a sequence of elements of $\mathcal{M}(X, V)$ such that μ_j takes values in a finite-dimensional linear subspace V_j of V for each j . Suppose also that $\{\mu_j\}_{j=1}^\infty$ converges to $\mu \in \mathcal{M}(X, V)$ with respect to the total variation norm, so that

$$(142.1) \quad \lim_{j \rightarrow \infty} \|\mu_j - \mu\|(X) = 0.$$

Let ρ be a finite nonnegative real measure on X such that $\|\mu_j\|$ is absolutely continuous with respect to ρ for each j , such as

$$(142.2) \quad \rho(A) = \sum_{j=1}^{\infty} a_j \|\mu_j\|(A)$$

for some $a_j > 0$ with $\sum_{j=1}^{\infty} a_j \|\mu_j\|(X) < \infty$. Thus each μ_j can be expressed as

$$(142.3) \quad \mu_j(A) = \int_A f_j d\rho$$

for some V_j -valued integrable function f_j on X with respect to ρ , by applying the Radon–Nikodym theorem to the components of $\mu(A)$ with respect to a basis for V_j as in the previous section. More precisely, each f_j is the sum of finitely many real or complex-valued integrable functions on X with respect to ρ times basis vectors of V_j , and the integral of f_j over A is the sum of the integrals of the components of f_j over A times the corresponding basis vectors of V_j . We also have that

$$(142.4) \quad \int_X \|f_j - f_l\| d\rho = \|\mu_j - \mu_l\|(X) \rightarrow 0$$

as $j, l \rightarrow \infty$, because of (142.1).

143 Uniform convexity

Let V be a vector space with a norm $\|v\|$. It will be convenient to take V to be a real vector space here, but complex vector spaces can also be considered as real vector spaces, and so everything in this section works as well in that case. We say that V is *uniformly convex* if for every $\epsilon > 0$ there is a $\delta > 0$ such that

$$(143.1) \quad v, w \in V, \quad \|v\| = \|w\| = 1, \quad \text{and} \quad \left\| \frac{v+w}{2} \right\| > 1 - \delta$$

imply that

$$(143.2) \quad \|v - w\| < \epsilon.$$

It is easy to see that inner product spaces are uniformly convex, because of the parallelogram law. It is well known that real and complex L^p spaces are uniformly convex when $1 < p < \infty$.

Suppose that $v, w \in V$, $\|v\|, \|w\| \leq 1$, and

$$(143.3) \quad \left\| \frac{v+w}{2} \right\| > 1 - \delta_1$$

for some $\delta_1 \in (0, 1/2)$. In particular,

$$(143.4) \quad \frac{\|v\| + \|w\|}{2} > 1 - \delta_1 > \frac{1}{2},$$

and so $\|v\|, \|w\| > 0$. If $v' = v/\|v\|$, $w' = w/\|w\|$, then

$$(143.5) \quad \|v' - v\| = (\|v\|^{-1} - 1) \|v\| = 1 - \|v\|,$$

and similarly for w . Thus

$$(143.6) \quad \frac{\|v' - v\| + \|w' - w\|}{2} = 1 - \frac{\|v\| + \|w\|}{2} < \delta_1,$$

which implies that

$$(143.7) \quad \begin{aligned} \left\| \frac{v+w}{2} \right\| &\leq \left\| \frac{v'+w'}{2} \right\| + \frac{\|v-v'\| + \|w-w'\|}{2} \\ &< \left\| \frac{v'+w'}{2} \right\| + \delta_1 \end{aligned}$$

and

$$(143.8) \quad \left\| \frac{v'+w'}{2} \right\| > 1 - 2\delta_1.$$

If δ_1 is sufficiently small, then

$$(143.9) \quad \|v' - w'\| < \epsilon/2,$$

by uniform convexity. If also $\delta_1 \leq \epsilon/4$, then

$$(143.10) \quad \|v - w\| \leq \|v' - w'\| + \|v - v'\| + \|w - w'\| < \epsilon/2 + 2\delta_1 \leq \epsilon.$$

This shows that uniform convexity implies the analogous condition in which $\|v\|, \|w\| \leq 1$.

Suppose that $v_1, \dots, v_n \in V$, $\|v_j\| \leq 1$ for $j = 1, \dots, n$, t_1, \dots, t_n are non-negative real numbers, and that $\sum_{j=1}^n t_j = 1$. Let $\epsilon > 0$ be given, and put

$$(143.11) \quad a = \sum_{j=1}^n t_j v_j.$$

Thus $\|a\| \leq 1$, and we would like to show that there is an $\eta > 0$ such that $\|a\| > 1 - \eta$ implies that

$$(143.12) \quad \sum_{j=1}^n t_j \|v_j - a\| < \epsilon,$$

where η does not depend on n , the v_j 's, or the t_j 's. Let λ be a bounded linear functional on V such that $\|\lambda\|_* = 1$ and $\lambda(a) = \|a\|$, the existence of which follows from the Hahn-Banach theorem, as usual. Hence

$$(143.13) \quad \sum_{j=1}^n t_j \lambda(v_j) = \lambda(a) = \|a\| > 1 - \eta,$$

which implies that

$$(143.14) \quad \sum_{j=1}^n t_j (1 - \lambda(v_j)) < \eta.$$

Note that $1 - \lambda(v_j) \geq 0$ for each j , because $|\lambda(v_j)| \leq 1$. In addition,

$$(143.15) \quad \left\| \frac{v_j + a}{2} \right\| \geq \lambda\left(\frac{v_j + a}{2}\right) = \frac{\lambda(v_j) + \|a\|}{2}.$$

Let δ_2 be associated to $\epsilon/2$ as in the second version of uniform convexity. If $\lambda(v_j) > 1 - \delta_2$ and $\eta \leq \delta_2$, then

$$(143.16) \quad \|(v_j + a)/2\| > \frac{(1 - \delta_2) + (1 - \eta)}{2} \geq 1 - \delta_2,$$

and so

$$(143.17) \quad \|v_j - a\| < \epsilon/2.$$

Let I_1 be the set of $j = 1, \dots, n$ such that $\lambda(v_j) > 1 - \delta_2$, and let I_2 be the set of $j = 1, \dots, n$ such that $\lambda(v_j) \leq 1 - \delta_2$. If $\eta \leq \delta_2$, then

$$(143.18) \quad \sum_{j \in I_1} t_j \|v_j - a\| < \epsilon/2,$$

by the preceding computation. Of course, $\|v_j - a\| \leq \|v_j\| + \|a\| \leq 2$ for each j , and so

$$(143.19) \quad \sum_{j \in I_2} t_j \|v_j - a\| \leq 2 \sum_{j \in I_2} t_j.$$

Using (143.14), we get that

$$(143.20) \quad \sum_{j \in I_2} t_j \delta_2 \leq \sum_{j \in I_2} t_j (1 - \lambda(v_j)) < \eta,$$

which implies that

$$(143.21) \quad \sum_{j \in I_2} t_j \|v_j - a\| \leq 2 \sum_{j \in I_2} t_j < 2 \delta_2^{-1} \eta.$$

Therefore

$$(143.22) \quad \begin{aligned} \sum_{j=1}^n t_j \|v_j - a\| &= \sum_{j \in I_1} t_j \|v_j - a\| + \sum_{j \in I_2} t_j \|v_j - a\| \\ &< \epsilon/2 + 2 \delta_2^{-1} \eta \leq \epsilon \end{aligned}$$

when η is sufficiently small, as desired.

144 Uniform convexity and measures

Let (X, \mathcal{A}) be a measurable space, and let $(V, \|v\|)$ be a uniformly convex Banach space. Also let $\epsilon > 0$ be given, and let η be as in the previous section. Suppose that $\mu \in \mathcal{M}(X, V)$ satisfies

$$(144.1) \quad \|\mu(X)\| > (1 - \eta) \|\mu\|(X).$$

Let $\mu_0 \in \mathcal{M}(X, V)$ be defined by

$$(144.2) \quad \mu_0(A) = \frac{\mu(X)}{\|\mu\|(X)} \|\mu\|(A),$$

so that μ_0 is the vector $\mu(X)/\|\mu\|(X)$ times the nonnegative real measure $\|\mu\|$ on X . We would like to show that

$$(144.3) \quad \|\mu - \mu_0\|(X) \leq \epsilon \|\mu\|(X)$$

under these conditions.

We may as well suppose also that $\|\mu\|(X) = 1$, since otherwise we can divide μ by $\|\mu\|(X) > 0$. Let A_1, \dots, A_n be finitely many pairwise disjoint measurable subsets of X such that $X = \bigcup_{j=1}^n A_j$, and let us check that

$$(144.4) \quad \sum_{j=1}^n \|\mu(A_j) - \mu_0(A_j)\| < \epsilon.$$

If $\|\mu\|(A_j) = 0$ for some j , then $\mu(A_j) = \mu_0(A_j) = 0$, and we can absorb A_j into one of the other A_l 's without affecting the sum. Thus we may as well ask that $\|\mu\|(A_j) > 0$ for each j too. If we put

$$(144.5) \quad t_j = \|\mu\|(A_j) \quad \text{and} \quad v_j = \frac{\mu(A_j)}{\|\mu\|(A_j)},$$

then $\sum_{j=1}^n t_j = 1$ and $\|v_j\| \leq 1$ for each j , because $\|\mu(A_j)\| \leq \|\mu\|(A_j)$. Also,

$$(144.6) \quad \sum_{j=1}^n t_j v_j = \sum_{j=1}^n \mu(A_j) = \mu(X),$$

and

$$(144.7) \quad \sum_{j=1}^n \|\mu(A_j) - \mu_0(A_j)\| = \sum_{j=1}^n t_j \|v_j - \mu(X)\|.$$

Thus (144.4) reduces to (143.12), with $a = \mu(X)$.

Now let μ be any element of $\mathcal{M}(X, V)$, and let θ be a small positive real number. By the definition of $\|\mu\|(X)$, there are finitely many pairwise-disjoint measurable sets X_1, \dots, X_r such that $X = \bigcup_{l=1}^r X_l$ and

$$(144.8) \quad \|\mu\|(X) < \sum_{l=1}^r \|\mu(X_l)\| + \theta.$$

Of course, $\|\mu\|(X) = \sum_{l=1}^r \|\mu\|(X_l)$, and so

$$(144.9) \quad \sum_{l=1}^r (\|\mu\|(X_l) - \|\mu(X_l)\|) < \theta.$$

Each term in the sum is nonnegative, since $\|\mu(X_l)\| \leq \|\mu\|(X_l)$. If L_2 is the set of $l = 1, \dots, r$ such that

$$(144.10) \quad \|\mu(X_l)\| \leq (1 - \eta) \|\mu\|(X_l),$$

where $\eta > 0$ is as before, then it follows that

$$(144.11) \quad \eta \sum_{l \in L_2} \|\mu\|(X_l) \leq \sum_{l \in L_2} (\|\mu\|(X_l) - \|\mu(X_l)\|) < \theta.$$

Let L_1 be the set of $l = 1, \dots, r$ such that $\|\mu(X_l)\| > (1 - \eta) \|\mu\|(X_l)$, and for each $l \in L_1$, let $\mu_l \in \mathcal{M}(X, V)$ be defined by

$$(144.12) \quad \mu_l(A) = \frac{\mu(X_l)}{\|\mu\|(X_l)} \|\mu\|(A \cap X_l).$$

This is analogous to (144.2), applied to the restriction of μ to X_l , and it follows from the earlier discussion that

$$(144.13) \quad \|\mu - \mu_l\|(X_l) \leq \epsilon \|\mu\|(X_l)$$

for each $l \in L_1$. Combining this with the earlier estimate (144.11) for L_2 , we get that

$$(144.14) \quad \left\| \mu - \sum_{l \in L_1} \mu_l \right\| < \epsilon \sum_{l \in L_1} \|\mu\|(X_l) + \eta^{-1} \theta \leq \epsilon \|\mu\|(X) + \eta^{-1} \theta.$$

Remember that η depends on ϵ , while θ can be chosen independently of ϵ , η . Thus the right side can be made arbitrarily small, by choosing ϵ and then θ appropriately.

145 Uniform convexity and paths

Let $(V, \|v\|)$ be a uniformly convex Banach space, and let $f : [x, y] \rightarrow V$ be a path of finite length Λ_x^y . Also let $\epsilon > 0$ be given, and let $\eta = \eta(\epsilon)$ be as in Section 143. Suppose that

$$(145.1) \quad \|f(x) - f(y)\| > (1 - \eta) \Lambda_x^y.$$

Put

$$(145.2) \quad f_0(z) = \frac{f(y) - f(x)}{\Lambda_x^y} \Lambda_x^z,$$

where Λ_x^z is the length of f on $[x, z]$, $x \leq z \leq y$. We would like to show that

$$(145.3) \quad \text{the length of } f - f_0 \text{ on } [x, y] \text{ is } \leq \epsilon \Lambda_x^y.$$

This is basically the same as the argument for measures in the previous section. As before, we may as well suppose that $\Lambda_x^y = 1$, since otherwise we can divide f by Λ_x^y .

If $\{r_j\}_{j=0}^n$ is any partition of $[x, y]$, then we would like to show that

$$(145.4) \quad \begin{aligned} & \sum_{j=1}^n \|(f(r_j) - f_0(r_j)) - (f(r_{j-1}) - f_0(r_{j-1}))\| \\ &= \sum_{j=1}^n \|(f(r_j) - f(r_{j-1})) - (f_0(r_j) - f_0(r_{j-1}))\| < \epsilon. \end{aligned}$$

We may as well ask that the length $\Lambda_{r_{j-1}}^{r_j}$ of f on $[r_{j-1}, r_j]$ be positive for each $j = 1, \dots, n$, since otherwise f, f_0 are constant on $[r_{j-1}, r_j]$, and r_j or r_{j-1} could be removed from the partition without affecting the sum. Put

$$(145.5) \quad t_j = \Lambda_{r_{j-1}}^{r_j} \quad \text{and} \quad v_j = \frac{f(r_j) - f(r_{j-1})}{\Lambda_{r_{j-1}}^{r_j}},$$

so that $\sum_{j=1}^n t_j = 1$ and $\|v_j\| \leq 1$ for each j , because $\|f(r_j) - f(r_{j-1})\| \leq \Lambda_{r_{j-1}}^{r_j}$. Moreover,

$$(145.6) \quad \sum_{j=1}^n t_j v_j = \sum_{j=1}^n (f(r_j) - f(r_{j-1})) = f(y) - f(x)$$

and

$$\begin{aligned}
 (145.7) \quad & \sum_{j=1}^n \|(f(r_j) - f(r_{j-1})) - (f_0(r_j) - f_0(r_{j-1}))\| \\
 &= \sum_{j=1}^n t_j \|v_j - (f(y) - f(x))\|.
 \end{aligned}$$

Thus (145.4) follows from (143.12), with $a = f(y) - f(x)$.

Now let $f : [a, b] \rightarrow V$ be a path of finite length Λ_a^b , and let θ be a small positive real number. By the definition of Λ_a^b , there is a partition $\{x_l\}_{l=0}^r$ of $[a, b]$ such that

$$(145.8) \quad \Lambda_a^b < \sum_{j=1}^r \|f(x_l) - f(x_{l-1})\| + \theta.$$

This implies that

$$(145.9) \quad \sum_{l=1}^r (\Lambda_{x_{l-1}}^{x_l} - \|f(x_l) - f(x_{l-1})\|) < \theta,$$

because $\Lambda_a^b = \sum_{l=1}^r \Lambda_{x_{l-1}}^{x_l}$. Note that the terms in the sum are nonnegative, since $\|f(x_l) - f(x_{l-1})\| \leq \Lambda_{x_{l-1}}^{x_l}$. If L_2 is the set of $l = 1, \dots, r$ such that

$$(145.10) \quad \|f(x_l) - f(x_{l-1})\| \leq (1 - \eta) \Lambda_{x_{l-1}}^{x_l},$$

where $\eta > 0$ is as before, then

$$(145.11) \quad \eta \sum_{l \in L_2} \Lambda_{x_{l-1}}^{x_l} \leq \sum_{l \in L_2} (\Lambda_{x_{l-1}}^{x_l} - \|f(x_l) - f(x_{l-1})\|) < \theta.$$

Let L_1 be the set of $l = 1, \dots, r$ such that

$$(145.12) \quad \|f(x_l) - f(x_{l-1})\| > (1 - \eta) \Lambda_{x_{l-1}}^{x_l}.$$

If $l \in L_2$, the define $f_l : [a, b] \rightarrow V$ by

$$(145.13) \quad f_l(z) = \frac{f(x_l) - f(x_{l-1})}{\Lambda_{x_{l-1}}^{x_l}} \Lambda_{x_{l-1}}^z$$

when $x_{l-1} \leq z \leq x_l$, and put $f_l(z) = 0$ when $z \leq x_{l-1}$, $f_l(z) = f(x_l) - f(x_{l-1})$ when $z \geq x_l$. This is the same as (145.2) on $[x_{l-1}, x_l]$ with $x = x_{l-1}$, $y = x_l$. As in (145.3), the length of $f - f_l$ on $[x_{l-1}, x_l]$ is less than or equal to $\epsilon \Lambda_{x_{l-1}}^{x_l}$. Combining this with (145.11), we get that the length of $f - \sum_{j \in L_1} f_j$ on $[a, b]$ is less than or equal to

$$(145.14) \quad \sum_{l \in L_1} \epsilon \Lambda_{x_{l-1}}^{x_l} + \eta^{-1} \theta \leq \epsilon \Lambda_a^b + \eta^{-1} \theta.$$

This uses the fact that the length of a path on $[a, b]$ is the sums of the lengths of its restrictions to the intervals $[x_{l-1}, x_l]$, $1 \leq l \leq r$. If $l \in L_1$, then f_j is constant on $[x_{l-1}, x_l]$ when $j \neq l$, by construction, and so the length of $f - \sum_{j \in L_1} f_j$ is the same as the length of $f - f_l$ on this interval. Similarly, if $l \in L_2$, then f_j is constant on $[x_{l-1}, x_l]$ for each $j \in L_1$, and the length of $f - \sum_{j \in L_1} f_j$ is the same as the length of f on this interval. It follows from this estimate that the length of $f - \sum_{j \in L_1} f_j$ can be made arbitrarily small, first by choosing ϵ to be very small, and then choosing θ to be sufficiently small, depending on η , which also depends on ϵ .

146 Uniform convexity and martingales

Let $(V, \|\cdot\|)$ be a uniformly convex Banach space. Also let $\epsilon > 0$ be given, and let $\eta > 0$ be as in Section 143. We may as well ask that $\eta \leq \epsilon$ too, which is practically unavoidable anyway.

Let (X, \mathcal{A}, μ) be a probability space, and let $\mathcal{P}_1, \mathcal{P}_2, \dots$ be a sequence of partitions of X into finitely or countably many pairwise disjoint measurable subsets of positive measure such that \mathcal{P}_{j+1} is a refinement of \mathcal{P}_j for each j . As usual, the arguments that follows are a bit simpler when each \mathcal{P}_j has only finitely many elements, but countable partitions and other situations can be accommodated as well. Let $\mathcal{B}_j = \mathcal{B}(\mathcal{P}_j)$ be the σ -algebra of measurable subsets of X generated by \mathcal{P}_j , as in Section 77, so that $\mathcal{B}_j \subseteq \mathcal{B}_{j+1}$ for each j .

We would like to consider V -valued martingales on X with respect to this filtration, as in Section 103. Remember that a V -valued function f_j on X is measurable with respect to \mathcal{B}_j if and only if it is constant on the elements of \mathcal{P}_j . Suppose that we have a sequence $\{f_j\}_{j=1}^\infty$ of V -valued functions on X such that f_j is measurable with respect to \mathcal{B}_j for each j and $\|f_j\|$ has bounded L^1 norm. Suppose also that $\{f_j\}_{j=1}^\infty$ is a martingale with respect to the \mathcal{B}_j 's, so that the value of f_j on $B \in \mathcal{P}_j$ is equal to the average of the values of f_{j+1} on the sets $A \in \mathcal{P}_{j+1}$ with $A \subseteq B$.

Under these conditions, $\{\|f_j\|\}_{j=1}^\infty$ is a submartingale on X with respect to the \mathcal{B}_j 's. In particular, the L^1 norm of $\|f_j\|$ is monotone increasing in j , and so

$$(146.1) \quad \lim_{j \rightarrow \infty} \int_X \|f_j\| d\mu = \sup_{j \geq 1} \int_X \|f_j\| d\mu.$$

Let θ be a small positive real number, and suppose that

$$(146.2) \quad \sup_{n \geq 1} \int_X \|f_n\| d\mu < \int_X \|f_j\| d\mu + \theta.$$

Note that

$$(146.3) \quad \int_B \|f_{n+1}\| d\mu \geq \int_B \|f_n\| d\mu$$

when $B \in \mathcal{P}_j$ and $n \geq j$, because $\{\|f_n\|\}_{n=1}^\infty$ is a submartingale. Moreover,

$$(146.4) \quad \lim_{n \rightarrow \infty} \int_X \|f_n\| d\mu = \lim_{n \rightarrow \infty} \sum_{B \in \mathcal{P}_j} \int_B \|f_n\| d\mu$$

$$= \sum_{B \in \mathcal{P}_j} \lim_{n \rightarrow \infty} \int_B \|f_n\| d\mu.$$

This is obvious when \mathcal{P}_j has only finitely many elements, and otherwise one can use the monotone convergence theorem for sums. It follows that

$$(146.5) \quad \sum_{B \in \mathcal{P}_j} \left(\lim_{n \rightarrow \infty} \int_B \|f_n\| d\mu - \int_B \|f_j\| d\mu \right) < \theta,$$

where each term in the sum is nonnegative.

Let \mathcal{P}'_j be the set of $B \in \mathcal{P}_j$ such that

$$(146.6) \quad \int_B \|f_j\| d\mu > (1 - \eta) \lim_{n \rightarrow \infty} \int_B \|f_n\| d\mu.$$

Thus $\mathcal{P}''_j = \mathcal{P}_j \setminus \mathcal{P}'_j$ consists of $B \in \mathcal{P}_j$ such that

$$(146.7) \quad \int_B \|f_j\| d\mu \leq (1 - \eta) \lim_{n \rightarrow \infty} \int_B \|f_n\| d\mu,$$

and satisfies

$$(146.8) \quad \eta \sum_{B \in \mathcal{P}''_j} \lim_{n \rightarrow \infty} \int_B \|f_n\| d\mu < \theta,$$

by (146.5).

Let $f_n(A)$ be the value of f_n on $A \in \mathcal{P}_n$, as in Section 103. Thus

$$(146.9) \quad f_j(B) = \sum_{\substack{A \in \mathcal{P}_n \\ A \subseteq B}} f_n(A) \frac{\mu(A)}{\mu(B)}$$

when $B \in \mathcal{P}_j$ and $n \geq j$, because $\{f_n\}_{n=1}^\infty$ is a martingale. In addition,

$$(146.10) \quad \int_B \|f_j\| d\mu = \|f_j(B)\| \mu(B)$$

and

$$(146.11) \quad \int_B \|f_n\| d\mu = \sum_{\substack{A \in \mathcal{P}_n \\ A \subseteq B}} \|f_n(A)\| \mu(A).$$

Note that

$$(146.12) \quad \int_B \|f_n\| d\mu \geq \int_B \|f_j\| d\mu > 0$$

when $B \in \mathcal{P}'_j$ and $n \geq j$, and put

$$(146.13) \quad t_n(A) = \|f_n(A)\| \mu(A) \left(\int_B \|f_n\| d\mu \right)^{-1}$$

for each $A \in \mathcal{P}_n$ with $A \subseteq B$, so that

$$(146.14) \quad \sum_{\substack{A \in \mathcal{P}_n \\ A \subseteq B}} t_n(A) = 1,$$

by construction. Also put $v_n(A) = f_n(A)/\|f_n(A)\|$ when $A \in \mathcal{P}_n$ and $f_n(A) \neq 0$, and $v_n(A) = 0$ when $f_n(A) = 0$, so that

$$(146.15) \quad \begin{aligned} \sum_{\substack{A \in \mathcal{P}_n \\ A \subseteq B}} v_n(A) t_n(A) &= \sum_{\substack{A \in \mathcal{P}_n \\ A \subseteq B}} f_n(A) \mu(A) \left(\int_B \|f_n\| d\mu \right)^{-1} \\ &= f_j(B) \mu(B) \left(\int_B \|f_n\| d\mu \right)^{-1}. \end{aligned}$$

It follows that

$$(146.16) \quad \left\| \sum_{\substack{A \in \mathcal{P}_n \\ A \subseteq B}} v_n(A) t_n(A) \right\| > 1 - \eta$$

when $B \in \mathcal{P}'_j$ and $n \geq j$.

This is exactly the situation discussed in Section 143, except that the sum in (146.16) may have infinitely many terms, which can be handled in the same way as before. If

$$(146.17) \quad a_{j,n}(B) = f_j(B) \mu(B) \left(\int_B \|f_n\| d\mu \right)^{-1},$$

then we get that

$$(146.18) \quad \sum_{\substack{A \in \mathcal{P}_n \\ A \subseteq B}} \|v_n(A) - a_{j,n}(B)\| t_n(A) < \epsilon$$

when $B \in \mathcal{P}'_j$ and $n \geq j$. Put $a_j(B) = f_j(B)/\|f_j(B)\|$, which is the same as $a_{j,j}(B)$, and observe that

$$(146.19) \quad a_{j,n}(B) = a_j(B) \frac{\int_B \|f_j\| d\mu}{\int_B \|f_n\| d\mu}.$$

This implies that

$$(146.20) \quad \|a_j(B) - a_{j,n}(B)\| < \eta$$

when $B \in \mathcal{P}'_j$ and $n \geq j$. Combining this with (146.18), we get that

$$(146.21) \quad \sum_{\substack{A \in \mathcal{P}_n \\ A \subseteq B}} \|v_n(A) - a_j(B)\| t_n(A) < \epsilon + \eta \leq 2\epsilon$$

when $B \in \mathcal{P}'_j$ and $n \geq j$.

Equivalently,

$$(146.22) \quad \sum_{\substack{A \in \mathcal{P}_n \\ A \subseteq B}} \|v_n(A) - a_j(B)\| \|f_n(A)\| \mu(A) < 2\epsilon \int_B \|f_n\| d\mu$$

when $B \in \mathcal{P}'_j$ and $n \geq j$, which reduces to

$$(146.23) \quad \sum_{\substack{A \in \mathcal{P}_n \\ A \subseteq B}} \|f_n(A) - a_j(B) \|f_n(A)\| \| \mu(A) < 2\epsilon \int_B \|f_n\| d\mu,$$

using the definition of $v_n(A)$. The sum on the left can be expressed as in integral, so that

$$(146.24) \quad \int_B \|f_n - a_j(B) \|f_n\| \| d\mu < 2\epsilon \int_B \|f_n\| d\mu$$

when $B \in \mathcal{P}'_j$ and $n \geq j$. Put $a_j(B) = 0$ when $B \in \mathcal{P}''_j$, and let $a_j(x)$ be the V -valued function on X equal to $a_j(B)$ when $x \in B \in \mathcal{P}$. Summing the previous estimate over $B \in \mathcal{P}'_j$, and using (146.8) for $B \in \mathcal{P}''_j$, we get that

$$(146.25) \quad \int_X \|f_n - a_j \|f_n\| \| d\mu < 2\epsilon \int_X \|f_n\| d\mu + \eta^{-1} \theta$$

when $n \geq j$.

As usual, the right side of (146.25) can be made arbitrarily small, by first choosing ϵ to be as small as one likes, and then choosing θ depending on η , which depends on ϵ . This works uniformly over $n \geq j$, because the L^1 norm of $\|f_n\|$ is bounded, by hypothesis. Because $\{\|f_n\|\}_{n=1}^\infty$ is a submartingale on X with bounded integral, there is a real-valued martingale $\{g_n\}_{n=1}^\infty$ on X such that $\|f_n\| \leq g_n$ and

$$(146.26) \quad \int_X g_n d\mu = \lim_{l \rightarrow \infty} \int_X \|f_l\| d\mu,$$

for each n , as in Section 91. Of course, the integral of g_n over X is independent of n , because of the martingale condition. In particular,

$$(146.27) \quad \int_X (g_n - \|f_n\|) d\mu < \theta$$

when $n \geq j$, by (146.2) and the monotonicity of the integral of $\|f_n\|$. Using (146.25), we get that

$$(146.28) \quad \int_X \|f_n - a_j g_n\| d\mu < 2\epsilon \int_X \|f_n\| d\mu + (\eta^{-1} + 1) \theta$$

when $n \geq j$, since $\|a_j(x)\| \leq 1$ for every $x \in X$, by construction. Note that $\{a_j g_n\}_{n=j}^\infty$ is a V -valued martingale on X , because $\{g_n\}_{n=1}^\infty$ is a martingale on X and a_j is constant on the elements of \mathcal{P}_j .

147 Strict convexity

Let V be a real vector space with a norm $\|v\|$. As before, a complex vector space is automatically a real vector space too, and so everything in this section can be used in that case as well. The closed unit ball

$$(147.1) \quad B_1 = \{v \in V : \|v\| \leq 1\}$$

in V is said to be *strictly convex* if for every $v, w \in B_1$ with $v \neq w$ and every $t \in \mathbf{R}$ with $0 < t < 1$ we have that

$$(147.2) \quad \|tv + (1-t)w\| < 1.$$

Of course, (147.2) holds automatically when $\|v\| < 1$ or $\|w\| < 1$, and so it suffices to check this when $\|v\| = \|w\| = 1$. One can show that the unit ball in an inner product space is strictly convex by determining when equality occurs in the Cauchy-Schwarz inequality. The unit ball in an L^p space is strictly convex when $1 < p < \infty$, because of the strict convexity of the function $|x|^p$ on the real line. This is similar to the proof of the convexity of the unit ball in L^p using the convexity of $|x|^p$, as in Section 8. Note that B_1 is strictly convex when V is uniformly convex.

Let λ be a nonzero bounded linear functional on V , and suppose that v, w are vectors in V such that $\|v\| = \|w\| = 1$ and $\lambda(v) = \lambda(w) = \|\lambda\|_*$. Thus

$$(147.3) \quad \lambda(tv + (1-t)w) = t\lambda(v) + (1-t)\lambda(w) = \|\lambda\|_*$$

when $0 < t < 1$, and hence

$$(147.4) \quad \|\lambda\|_* = |\lambda(tv + (1-t)w)| \leq \|\lambda\|_* \|tv + (1-t)w\|,$$

which implies that

$$(147.5) \quad \|tv + (1-t)w\| \geq 1.$$

By the triangle inequality, $\|tv + (1-t)w\| \leq 1$ when $0 < t < 1$, and so

$$(147.6) \quad \|tv + (1-t)w\| = 1.$$

If B_1 is strictly convex, then it follows that $v = w$ under these conditions. Conversely, let us check that this property characterizes strict convexity of B_1 .

Suppose that $v, w \in V$, $\|v\| = \|w\| = 1$, $0 < t < 1$, and that $a = tv + (1-t)w$ satisfies $\|a\| = 1$. As usual, there is a bounded linear functional λ on V such that $\lambda(a) = \|\lambda\|_* = 1$, because of the Hahn-Banach theorem. This implies that $|\lambda(v)|, |\lambda(w)| \leq 1$ and

$$(147.7) \quad 1 = \lambda(tv + (1-t)w) = t\lambda(v) + (1-t)\lambda(w),$$

so that $\lambda(v) = \lambda(w) = 1$. If we have the uniqueness property described in the previous paragraph, then we get that $v = w$, which means that B_1 is strictly convex.

If V is not uniformly convex, then there is an $\epsilon > 0$ and sequences of vectors $\{v_j\}_{j=1}^\infty, \{w_j\}_{j=1}^\infty$ in V such that $\|v_j\| = \|w_j\| = 1$ and $\|v_j - w_j\| \geq \epsilon$ for each j , and

$$(147.8) \quad \lim_{j \rightarrow \infty} \left\| \frac{v_j + w_j}{2} \right\| = 1.$$

If V has finite dimension n , then there is a one-to-one linear mapping from \mathbf{R}^n onto V . This mapping is also a homeomorphism with respect to the standard

topology on \mathbf{R}^n and the topology on V determined by the metric associated to the norm. In particular, closed and bounded subsets of V are compact in this case. Thus we may suppose in addition that $\{v_j\}_{j=1}^\infty, \{w_j\}_{j=1}^\infty$ converge to some vectors $v, w \in V$, respectively, by passing to subsequences. By hypothesis, $\|v\| = \|w\| = 1$, $\|v - w\| \geq \epsilon > 0$, and $\|(v + w)/2\| = 1$, which is impossible when B_1 is strictly convex. This shows that V is uniformly convex when V is finite-dimensional and B_1 is strictly convex.

Suppose that B_1 is strictly convex, and that

$$(147.9) \quad \|v + w\| = \|v\| + \|w\|$$

for some $v, w \in V$ with $v, w \neq 0$. If

$$(147.10) \quad v' = \frac{v}{\|v\|}, \quad w' = \frac{w}{\|w\|}, \quad \text{and } t = \frac{\|v\|}{\|v\| + \|w\|},$$

then $1 - t = \|w\|/(\|v\| + \|w\|)$ and

$$(147.11) \quad t v' + (1 - t) w' = \frac{v + w}{\|v\| + \|w\|}.$$

This has norm 1 by hypothesis, so that $v' = w'$ by strict convexity. Equivalently, $w = r v$, where $r = \|w\|/\|v\|$.

Let (X, \mathcal{A}) be a measurable space, and suppose that $\mu \in \mathcal{M}(X, V)$. If $A \subseteq X$ is measurable, then $\mu(X) = \mu(A) + \mu(X \setminus A)$, which implies that

$$(147.12) \quad \begin{aligned} \|\mu(X)\| &\leq \|\mu(A)\| + \|\mu(X \setminus A)\| \\ &\leq \|\mu\|(A) + \|\mu\|(X \setminus A) = \|\mu\|(X). \end{aligned}$$

If $\|\mu(X)\| = \|\mu\|(X)$, then it follows that

$$(147.13) \quad \|\mu(X)\| = \|\mu(A)\| + \|\mu(X \setminus A)\|$$

and

$$(147.14) \quad \|\mu(A)\| = \|\mu\|(A)$$

for every measurable set $A \subseteq X$. If $\|\mu\|(X) > 0$ and B_1 is strictly convex, then one can argue as in the preceding paragraph to get that

$$(147.15) \quad \mu(A) = \mu(X) \frac{\|\mu\|(A)}{\|\mu\|(X)}$$

for every measurable set $A \subseteq X$.

Suppose now that $f : [a, b] \rightarrow V$ is a path of finite length, and let Λ_x^y be the length of the restriction of f to $[x, y] \subseteq [a, b]$. Thus

$$(147.16) \quad \begin{aligned} \|f(b) - f(a)\| &\leq \|f(x) - f(a)\| + \|f(b) - f(x)\| \\ &\leq \Lambda_a^x + \Lambda_x^b = \Lambda_a^b \end{aligned}$$

when $a \leq x \leq b$. If $\|f(b) - f(a)\| = \Lambda_a^b$, then it follows that

$$(147.17) \quad \|f(b) - f(a)\| = \|f(x) - f(a)\| + \|f(b) - f(x)\|$$

and

$$(147.18) \quad \|f(x) - f(a)\| = \Lambda_a^x$$

when $a \leq x \leq b$. If $\Lambda_a^b > 0$ and B_1 is strictly convex, then one can argue as before to get that

$$(147.19) \quad f(x) - f(a) = (f(b) - f(a)) \frac{\Lambda_a^x}{\Lambda_a^b}$$

when $a \leq x \leq b$.

148 Minimizing distances

Let $(V, \|v\|)$ be a uniformly convex Banach space, and let E be a nonempty closed convex set in V . Also let $v \in V$ be given, and let ρ be the distance from v to E ,

$$(148.1) \quad \rho = \inf\{\|v - w\| : w \in E\}.$$

Let $\{w_j\}_{j=1}^\infty$ be a sequence of elements of E such that

$$(148.2) \quad \lim_{j \rightarrow \infty} \|v - w_j\| = \rho.$$

Because E is convex, $(w_j + w_l)/2 \in E$ for every $j, l \geq 1$, and so

$$(148.3) \quad \left\|v - \frac{w_j + w_l}{2}\right\| \geq \rho.$$

Suppose that $v \notin E$, so that $\rho > 0$, and put

$$(148.4) \quad u_j = \frac{v - w_j}{\|v - w_j\|}$$

for each j . Thus $\|u_j\| = 1$ for each j , and hence $\|(u_j + u_l)/2\| \leq 1$ for every $j, l \geq 1$, by the triangle inequality. Using (148.2) and (148.3), it is easy to see that

$$(148.5) \quad \lim_{j, l \rightarrow \infty} \left\|\frac{u_j + u_l}{2}\right\| = 1.$$

This implies that

$$(148.6) \quad \lim_{j, l \rightarrow \infty} \|u_j - u_l\| = 0,$$

because of uniform convexity. Using (148.2) again, it is easy to check that

$$(148.7) \quad \|w_j - w_l\| = \|(v - w_j) - (v - w_l)\| \rightarrow 0 \text{ as } j, l \rightarrow \infty.$$

This shows that $\{w_j\}_{j=1}^\infty$ is a Cauchy sequence, which therefore converges to some $w \in V$. We also have that $w \in E$, because E is closed. Of course, $\|v - w\| = \rho$, so that w minimizes the distance to v from elements of E .

Suppose that w' is another element of E such that $\|v - w'\| = \rho$. If $0 < t < 1$, then $tw + (1 - t)w' \in E$, because E is convex, and so

$$(148.8) \quad \|v - (tw + (1 - t)w')\| \geq \rho.$$

Moreover,

$$(148.9) \quad \|v - (tw + (1 - t)w')\| \leq t\|v - w\| + (1 - t)\|v - w'\| = \rho,$$

which implies that

$$(148.10) \quad \|v - (tw + (1 - t)w')\| = \rho.$$

Put $u = \rho^{-1}(v - w)$, $u' = \rho^{-1}(v - w')$, so that $\|u\| = \|u'\| = 1$ and

$$(148.11) \quad \|tu + (1 - t)u'\| = 1$$

when $0 < t < 1$. Strict convexity of the closed unit ball in V implies that $u = u'$, which is the same as saying that $w = w'$.

Let λ be a nonzero bounded linear functional on V , and let E be the set of $w \in V$ such that $\lambda(w) = \|\lambda\|_*$. This is a closed affine subspace of V , which is convex in particular. The distance ρ from E to 0 is the same as the infimum of $\|w\|$ over $w \in E$, which is equal to 1 in this case, by the definition of the dual norm of λ . The arguments in the previous paragraphs imply that there is a unique $w \in E$ such that $\|w\| = 1$. This shows that the supremum is attained in the definition of the dual norm of a bounded linear functional on a uniformly convex Banach space.

149 Another approximation argument

Let V_1 be a real vector space with a norm $\|v\|$. As usual, everything in this section can also be applied to complex vector spaces, since they are real vector spaces too. Suppose that V_1 is uniformly convex, so that for each $\epsilon > 0$ there is a $\delta(\epsilon) > 0$ such that for every $v, w \in V_1$ with $\|v\| = \|w\| = 1$ and

$$(149.1) \quad \left\| \frac{v + w}{2} \right\| > 1 - \delta(\epsilon)$$

we have that $\|v - w\| < \epsilon$, as in Section 143. Although uniform convexity follows from strict convexity of the unit ball in finite dimensions, as in Section 147, the estimates in this section will only depend on $\delta(\epsilon)$, and not on the particular norm $\|v\|$, or the dimension of V_1 . Hence these estimates hold uniformly over all finite-dimensional subspaces of a uniformly convex Banach space, for instance.

Let (X, \mathcal{A}, μ) be a probability space, and let \mathcal{B} be a σ -subalgebra of \mathcal{A} . As in Section 120, it is easy to deal with integrals of V_1 -valued functions on X , by integrating the components of these functions with respect to a basis for V_1 . Similarly, the conditional expectation of a V_1 -valued function on X with respect to \mathcal{B} can be defined by taking the conditional expectation of the components of

the function with respect to a basis. It is easy to see that this does not depend on the choice of a basis for V_1 , using the linearity of integration and conditional expectation.

Let f be an integrable V_1 -valued function on X with respect to μ , which means that the components of f with respect to a basis are integrable real-valued functions. Also let $f_{\mathcal{B}} = E(f \mid \mathcal{B})$ be the conditional expectation of f with respect to \mathcal{B} , as usual. Remember that

$$(149.2) \quad \|f_{\mathcal{B}}\| \leq E(\|f\| \mid \mathcal{B})$$

almost everywhere on X , as in Section 120, so that

$$(149.3) \quad \int_X \|f_{\mathcal{B}}\| d\mu \leq \int_X \|f\| d\mu,$$

in particular. Let θ be a small positive real number, and suppose that

$$(149.4) \quad \int_X \|f\| d\mu < \int_X \|f_{\mathcal{B}}\| d\mu + \theta.$$

This implies that

$$(149.5) \quad \int_X (E(\|f\| \mid \mathcal{B}) - \|f_{\mathcal{B}}\|) d\mu < \theta,$$

because the integrals of $\|f\|$ and $E(\|f\| \mid \mathcal{B})$ over X are the same, since $X \in \mathcal{B}$.

Let η be another small positive real number, and put

$$(149.6) \quad X_1 = \{x \in X : \|f_{\mathcal{B}}(x)\| > (1 - \eta) E(\|f\| \mid \mathcal{B})(x)\},$$

$$(149.7) \quad X_2 = \{x \in X : \|f_{\mathcal{B}}(x)\| \leq (1 - \eta) E(\|f\| \mid \mathcal{B})(x)\}.$$

Thus $X_1, X_2 \in \mathcal{B}$, because $\|f_{\mathcal{B}}\|$, $E(\|f\| \mid \mathcal{B})$ are measurable with respect to \mathcal{B} . Note that

$$(149.8) \quad \begin{aligned} \eta \int_{X_2} \|f\| d\mu &= \int_{X_2} \eta E(\|f\| \mid \mathcal{B}) d\mu \\ &\leq \int_X (E(\|f\| \mid \mathcal{B}) - \|f_{\mathcal{B}}\|) d\mu < \theta, \end{aligned}$$

where we use the fact that $X_2 \in \mathcal{B}$ in the first step, and (149.2) and the definition of X_2 in the second step.

In order to see what happens on X_1 , it will be convenient to use linear functionals on V_1 . Of course, every linear functional on V_1 is bounded, because V_1 has finite dimension, and the dual V_1^* of V_1 has finite dimension equal to the dimension of V_1 . In particular, there is a sequence of linear functionals $\{\lambda_j\}_{j=1}^{\infty}$ on V_1 such that $\|\lambda_j\|_* = 1$ for each j and the λ_j 's are dense in the set of $\lambda \in V_1^*$ with $\|\lambda\|_* = 1$. As usual, for each $v \in V_1$ there is a $\lambda \in V_1^*$ such that $\|\lambda\|_* = 1$ and $\lambda(v) = \|v\|$, because of the Hahn–Banach theorem. This implies that

$$(149.9) \quad \|v\| = \sup_{j \geq 1} \lambda_j(v)$$

for each $v \in V_1$, by approximating λ by λ_j 's, and using the fact that $\|\lambda_j\|_* = 1$ for each j .

Put

$$(149.10) \quad A_j = \{x \in X : \lambda_j(f_{\mathcal{B}}(x)) > (1 - \eta) E(\|f\| \mid \mathcal{B})(x)\}$$

for each $j \geq 1$, so that $A_j \subseteq X_1$ and $A_j \in \mathcal{B}$ for each j , and

$$(149.11) \quad \bigcup_{j=1}^{\infty} A_j = X_1,$$

by (149.9). It is better to have disjoint sets, and so we let $B_1 = A_1$ and $B_n = A_n \setminus \left(\bigcup_{j=1}^{n-1} A_j\right)$ when $n \geq 2$. Thus $B_n \subseteq A_n \subseteq X_1$ and $B_n \in \mathcal{B}$ for each n , $B_l \cap B_n = \emptyset$ when $l < n$, and

$$(149.12) \quad \bigcup_{n=1}^{\infty} B_n = \bigcup_{j=1}^{\infty} A_j = X_1,$$

as before. Note that $\lambda \circ f_{\mathcal{B}} = E(\lambda \circ f \mid \mathcal{B})$ for each linear functional λ on V_1 . This implies that

$$(149.13) \quad \int_{B_n} \lambda_n \circ f_{\mathcal{B}} d\mu = \int_{B_n} \lambda_n \circ f d\mu,$$

since $B_n \in \mathcal{B}$, while

$$(149.14) \quad \int_{B_n} E(\|f\| \mid \mathcal{B}) d\mu = \int_{B_n} \|f\| d\mu.$$

Because $B_n \subseteq A_n$,

$$(149.15) \quad \int_{B_n} \lambda_n \circ f_{\mathcal{B}} d\mu > (1 - \eta) \int_{B_n} E(\|f\| \mid \mathcal{B}) d\mu$$

when $\mu(B_n) > 0$, and hence

$$(149.16) \quad \int_{B_n} \lambda_n \circ f d\mu > (1 - \eta) \int_{B_n} \|f\| d\mu.$$

Equivalently,

$$(149.17) \quad \int_{B_n} (\|f\| - \lambda_n \circ f) d\mu < \eta \int_{B_n} \|f\| d\mu$$

when $\mu(B_n) > 0$, where the integrand on the left is nonnegative, since $\|\lambda_n\|_* = 1$.

Let $\epsilon > 0$ be given, and put $\delta = \delta(\epsilon)$. Also put

$$(149.18) \quad B_{n,1} = \{x \in B_n : \lambda_n(f(x)) > (1 - \delta) \|f(x)\|\},$$

$$(149.19) \quad B_{n,2} = \{x \in B_n : \lambda_n(f(x)) \leq (1 - \delta) \|f(x)\|\}.$$

Thus

$$(149.20) \quad \delta \int_{B_{n,2}} \|f\| d\mu \leq \int_{B_{n,2}} (\|f\| - \lambda_n \circ f) d\mu < \eta \int_{B_n} \|f\| d\mu$$

when $\mu(B_n) > 0$, by (149.17). As before, we shall be interested in η 's that are small compared to δ , so that the integral of $\|f\|$ over $B_{n,2}$ is relatively small.

If $x \in X_1$, then $f_{\mathcal{B}}(x) \neq 0$, and we put $a(x) = f_{\mathcal{B}}(x)/\|f_{\mathcal{B}}(x)\|$. Otherwise, if $x \in X_2$, then we put $a(x) = 0$. If $x \in B_n \subseteq A_n \subseteq X_1$, then

$$(149.21) \quad \lambda_n(a(x)) > 1 - \eta,$$

using also (149.2). If $x \in B_{n,1}$, then $f(x) \neq 0$, and we put $b(x) = f(x)/\|f(x)\|$. Note that

$$(149.22) \quad \lambda_n(b(x)) > 1 - \delta,$$

by definition of $B_{n,1}$. Thus $\|a(x)\| = \|b(x)\| = 1$ and

$$(149.23) \quad \left\| \frac{a(x) + b(x)}{2} \right\| \geq \lambda_n\left(\frac{a(x) + b(x)}{2}\right) > 1 - \frac{\delta + \eta}{2} \geq 1 - \delta$$

when $x \in B_{n,1}$ and $\eta \leq \delta$. This implies that $\|a(x) - b(x)\| < \epsilon$, because of uniform convexity. Equivalently,

$$(149.24) \quad \|f(x) - a(x)\| \|f(x)\| < \epsilon \|f(x)\|$$

when $x \in B_{n,1}$ and $\eta \leq \delta$.

It follows that

$$(149.25) \quad \int_{B_n} \|f(x) - a(x)\| \|f(x)\| d\mu \leq \int_{B_{n,1}} \epsilon \|f\| d\mu + \int_{B_{n,2}} 2 \|f\| d\mu$$

when $\eta \leq \delta(\epsilon)$, and hence

$$(149.26) \quad \int_{B_n} \|f(x) - a(x)\| \|f(x)\| d\mu \leq (\epsilon + 2\delta(\epsilon)^{-1}\eta) \int_{B_n} \|f\| d\mu,$$

because of (149.20). This also holds trivially when $\eta > \delta(\epsilon)$, since the coefficient on the right would be greater than 2. Summing over n , we get that

$$(149.27) \quad \int_{X_1} \|f(x) - a(x)\| \|f(x)\| d\mu \leq (\epsilon + 2\delta(\epsilon)^{-1}\eta) \int_{X_1} \|f\| d\mu.$$

Combining this with (149.8), we obtain

$$(149.28) \quad \int_X \|f(x) - a(x)\| \|f(x)\| d\mu < (\epsilon + 2\delta(\epsilon)^{-1}\eta) \int_X \|f\| d\mu + \eta^{-1}\theta.$$

Alternatively, one might prefer to take $a(x) = f_{\mathcal{B}}(x)/\|f_{\mathcal{B}}(x)\|$ for every x in X such that $f_{\mathcal{B}}(x) \neq 0$, even when $x \in X_2$. This would ensure that $a(x)$ does not depend on f even indirectly, through the definition of X_2 . In this case, we would get that

$$(149.29) \quad \int_X \|f(x) - a(x)\| \|f(x)\| d\mu < (\epsilon + 2\delta(\epsilon)^{-1}\eta) \int_X \|f\| d\mu + 2\eta^{-1}\theta,$$

which is to say that we would multiply $\eta^{-1}\theta$ by 2 in the previous estimate. In both situations, $a(x)$ is measurable with respect to \mathcal{B} , because $f_{\mathcal{B}}$ is measurable with respect to \mathcal{B} and $X_2 \in \mathcal{B}$.

150 Examples in ℓ^p

Let a_1, a_2, \dots be a sequence of real or complex numbers, and consider

$$(150.1) \quad f_n(x) = \sum_{j=1}^n a_j r_j(x) \delta_j.$$

Here $r_1(x), r_2(x), \dots$ are the Rademacher functions, and $\delta_j = \{\delta_{j,l}\}_{l=1}^\infty$ is the sequence defined by $\delta_{j,l} = 1$ when $j = l$ and $\delta_{j,l} = 0$ when $j \neq l$. Thus $\{f_n\}_{n=1}^\infty$ is a martingale on the dyadic unit interval with respect to the usual filtration associated to dyadic subintervals, and with values in the vector space of sequences of real or complex numbers, as appropriate. In particular, $\{f_n\}_{n=1}^\infty$ is a martingale with values in ℓ^p for each p , $1 \leq p \leq \infty$. Note that the ℓ^p norm of $f_n(x)$ is equal to the ℓ^p norm of the finite sequence a_1, \dots, a_n for each x and n . Hence the L^1 norm of $\|f_n(x)\|_{\ell^p}$ is equal to the ℓ^p norm of a_1, \dots, a_n for each n . It follows that the L^1 norm of $\|f_n(x)\|_{\ell^p}$ is uniformly bounded over n if and only if $\{a_j\}_{j=1}^\infty$ is in ℓ^p . If $\{a_j\}_{j=1}^\infty \in \ell^p$ and $p < \infty$, then it is easy to see that $f_n(x)$ converges in ℓ^p as $n \rightarrow \infty$ for each x . Similarly, if $\{a_j\}_{j=1}^\infty$ converges to 0, then $f_n(x)$ converges in c_0 equipped with the ℓ^∞ norm as $n \rightarrow \infty$ for each x . If $\{a_j\}_{j=1}^\infty$ is bounded, then $f_n(x)$ is uniformly bounded in ℓ^∞ , but it does not converge in the ℓ^∞ norm as $n \rightarrow \infty$ for any x unless $\{a_j\}_{j=1}^\infty$ converges to 0.

151 Uniform convergence

Let $(V, \|v\|)$ be a real or complex Banach space, and let v_1, v_2, \dots be a sequence of elements of V . As in Section 60, let X be the set of sequences $x = \{x_j\}_{j=1}^\infty$ with $x_j = 1$ or -1 for each j , which is the same as the Cartesian product of a sequence of copies of $\{1, -1\}$. Consider

$$(151.1) \quad f_n(x) = \sum_{j=1}^n x_j v_j$$

for each positive integer n and $x \in X$. This is basically the same as the sequence of functions considered in the previous section when $V = \ell^p$ and $v_j = a_j \delta_j$, since $r_j(x) = x_j$ is another version of the Rademacher functions. Let us check that $\{f_n\}_{n=1}^\infty$ converges uniformly on X when $\sum_{j \in \mathbf{Z}_+} v_j$ converges in the generalized sense, as in Section 14. In particular, $\{f_n\}_{n=1}^\infty$ converges uniformly on X when $\sum_{j=1}^\infty v_j$ converges absolutely. In this case, it is very easy to show directly that $\{f_n\}_{n=1}^\infty$ converges uniformly, by the same argument as in Weierstrass' M -test.

Suppose that $\sum_{j \in \mathbf{Z}_+} v_j$ converges in the generalized sense, which implies that it satisfies the generalized Cauchy criterion, as in Section 14. This means that for each $\epsilon > 0$ there is a finite set $A_\epsilon \subseteq \mathbf{Z}_+$ such that

$$(151.2) \quad \left\| \sum_{j \in B} v_j \right\| < \epsilon$$

for every finite set $B \subseteq \mathbf{Z}_+$ with $A_\epsilon \cap B = \emptyset$. Let L_ϵ be the maximum of the elements of A_ϵ , with $L_\epsilon = 0$ when $A_\epsilon = \emptyset$. If $n > l$, then

$$(151.3) \quad f_n(x) - f_l(x) = \sum_{j=l+1}^n x_j v_j = \sum_{j \in B_{l,n}^+} v_j - \sum_{j \in B_{l,n}^-} v_j,$$

where $B_{l,n}^+$, $B_{l,n}^-$ are the sets of positive integers j such that $l < j \leq n$ and $x_j = 1$ or -1 , respectively. If $l \geq L_\epsilon$, then $B_{l,n}^+ \cap A_\epsilon = B_{l,n}^- \cap A_\epsilon = \emptyset$, and so

$$(151.4) \quad \|f_n(x) - f_l(x)\| \leq \left\| \sum_{j \in B_{l,n}^+} v_j \right\| + \left\| \sum_{j \in B_{l,n}^-} v_j \right\| < \epsilon + \epsilon = 2\epsilon.$$

This shows that $\{f_n\}_{n=1}^\infty$ is a Cauchy sequence with respect to the supremum norm on the space of V -valued functions on X . It follows that $\{f_n\}_{n=1}^\infty$ converges uniformly on X , because V is complete. As usual, one can observe first that $\{f_n(x)\}_{n=1}^\infty$ is a Cauchy sequence in V for each $x \in X$, which converges because of completeness, and then check that $\{f_n\}_{n=1}^\infty$ converges uniformly on X to the pointwise limit, because of the uniform version of the Cauchy condition.

Conversely, suppose that $\{f_n\}_{n=1}^\infty$ converges uniformly on X , and hence satisfies the uniform version of the Cauchy condition. This means that for each $\epsilon > 0$ there is an $N_\epsilon \geq 0$ such that

$$(151.5) \quad \|f_n(x) - f_l(x)\| < \epsilon$$

for every $n > l \geq N_\epsilon$ and $x \in X$, or equivalently

$$(151.6) \quad \left\| \sum_{j=l+1}^n x_j v_j \right\| < \epsilon$$

for every $n > l \geq N_\epsilon$ and $x \in X$. Let $B \subseteq \mathbf{Z}_+$ be a nonempty finite set whose minimal element is greater than N_ϵ . If $y, z \in X$ are defined by $y_j = 1$ for every j , $z_j = 1$ when $j \in B$, and $z_j = -1$ otherwise, then

$$(151.7) \quad \sum_{j=N_\epsilon+1}^n y_j v_j + \sum_{j=N_\epsilon+1}^n z_j v_j = 2 \sum_{j \in B} v_j$$

when the maximal element of B is less than or equal to n . Hence

$$(151.8) \quad 2 \left\| \sum_{j \in B} v_j \right\| \leq \left\| \sum_{j=N_\epsilon+1}^n y_j v_j \right\| + \left\| \sum_{j=N_\epsilon+1}^n z_j v_j \right\| < \epsilon + \epsilon = 2\epsilon,$$

by (151.6). This is the same as saying that $\left\| \sum_{j \in B} v_j \right\| < \epsilon$ when $B \subseteq \mathbf{Z}_+$ is a finite set disjoint from $\{1, \dots, N_\epsilon\}$, which implies that $\sum_{j \in \mathbf{Z}_+} v_j$ satisfies the generalized Cauchy criterion. Thus $\sum_{j \in \mathbf{Z}_+} v_j$ converges in the generalized sense, because V is complete.

Actually, the same conclusion holds when $\{f_n(x)\}_{n=1}^\infty$ converges in V for every $x \in X$, which is the same as saying that $\sum_{j=1}^\infty x_j v_j$ converges for every $x \in X$. To see this, suppose for the sake of a contradiction that $\sum_{j \in \mathbf{Z}_+} v_j$ does not satisfy the generalized Cauchy condition. This means that for each $\epsilon > 0$ and finite set $A \subseteq \mathbf{Z}_+$ there is a finite set $B \subseteq \mathbf{Z}_+$ such that

$$(151.9) \quad \left\| \sum_{j \in B} v_j \right\| \geq \epsilon.$$

By applying this repeatedly, we can get an infinite sequence B_1, B_2, \dots of finite subsets of \mathbf{Z}_+ such that the maximal element of B_l is strictly less than the minimal element of B_{l+1} for each l , and (151.9) holds with $B = B_l$ for each l . Let $y, z \in X$ be defined by $y_j = 1$ for each j , $z_j = 1$ when $j \in B_l$ for some $l \geq 1$, and $z_j = -1$ otherwise. If b_n is the maximal element of B_n , then

$$(151.10) \quad \sum_{j=1}^{b_n} y_j v_j + \sum_{j=1}^{b_n} z_j v_j = 2 \sum_{l=1}^n \left(\sum_{j \in B_l} v_j \right).$$

Thus the convergence of $\sum_{j=1}^\infty y_j v_j$ and $\sum_{j=1}^\infty z_j v_j$ imply the convergence of

$$(151.11) \quad \sum_{l=1}^\infty \left(\sum_{j \in B_l} v_j \right).$$

This implies in turn that

$$(151.12) \quad \lim_{l \rightarrow \infty} \sum_{j \in B_l} v_j = 0,$$

a contradiction. This shows that $\sum_{j \in \mathbf{Z}_+} v_j$ satisfies the generalized Cauchy condition, and hence converges in the generalized sense, because V is complete. Therefore $\sum_{j \in \mathbf{Z}_+} v_j$ converges in the generalized sense if and only if $\sum_{j=1}^\infty x_j v_j$ converges for every $x \in X$, in which case the partial sums f_n converge uniformly on X .

152 Bounded sums

Let V be a real or complex vector space with a norm $\|v\|$, and let $\{1, -1\}^n$ be the Cartesian product of n copies of $\{1, -1\}$, consisting of all sequences $\epsilon = \{\epsilon_j\}_{j=1}^n$ of length n with $\epsilon_j = 1$ or -1 for each j . Also let $Z(V)$ be the collection of sequences v_1, v_2, \dots of vectors in V for which the sums $\sum_{j=1}^n \epsilon_j v_j$ are uniformly bounded in V over $\epsilon \in \{1, -1\}^n$ and all positive integers n . This is a vector space over the real or complex numbers, as appropriate, with respect to termwise addition and scalar multiplication. If $\{v_j\}_{j=1}^\infty \in Z(V)$, then put

$$(152.1) \quad \|\{v_j\}_{j=1}^\infty\|_{Z(V)} = \sup \left\{ \left\| \sum_{j=1}^n \epsilon_j v_j \right\| : \epsilon \in \{1, -1\}^n, n \in \mathbf{Z}_+ \right\}.$$

Note that $Z(V)$ is a linear subspace of the space $X(V)$ of sequences $\{v_j\}_{j=1}^\infty$ of vectors in V with bounded partial sums $\sum_{j=1}^n v_j$, discussed in Section 26, since we can take $\epsilon_j = 1$ for each j . Similarly,

$$(152.2) \quad \|\{v_j\}_{j=1}^\infty\|_{X(V)} \leq \|\{v_j\}_{j=1}^\infty\|_{Z(V)}$$

for each $\{v_j\}_{j=1}^\infty \in Z(V)$. It is easy to see that $\|\{v_j\}_{j=1}^\infty\|_{Z(V)}$ is a norm on $Z(V)$, and in particular that $v_j = 0$ for every j when $\|\{v_j\}_{j=1}^\infty\|_{Z(V)} = 0$.

Let $\{v_j\}_{j=1}^\infty$ be a sequence of vectors in V , let B be a finite nonempty set of positive integers, and let n be the maximal element of B . If $\alpha, \beta \in \{1, -1\}^n$ are defined by $\alpha_j = 1$ for each j , $\beta_j = 1$ when $j \in B$, and $\beta_j = -1$ otherwise, then

$$(152.3) \quad \sum_{j=1}^n \alpha_j v_j + \sum_{j=1}^n \beta_j v_j = 2 \sum_{j \in B} v_j,$$

and hence

$$(152.4) \quad 2 \left\| \sum_{j \in B} v_j \right\| \leq \left\| \sum_{j=1}^n \alpha_j v_j \right\| + \left\| \sum_{j=1}^n \beta_j v_j \right\|.$$

If $\{v_j\}_{j=1}^\infty \in Z(V)$, then we get that

$$(152.5) \quad \left\| \sum_{j \in B} v_j \right\| \leq \|\{v_j\}_{j=1}^\infty\|_{Z(V)},$$

which implies that $\{v_j\}_{j=1}^\infty$ is in the space $Y(\mathbf{Z}_+, V)$ discussed in Section 27, and that

$$(152.6) \quad \|\{v_j\}_{j=1}^\infty\|_{Y(\mathbf{Z}_+, V)} \leq \|\{v_j\}_{j=1}^\infty\|_{Z(V)}.$$

Conversely, if $\{v_j\}_{j=1}^\infty \in Y(\mathbf{Z}_+, V)$, $n \in \mathbf{Z}_+$, and $\epsilon \in \{1, -1\}^n$, then

$$(152.7) \quad \sum_{j=1}^n \epsilon_j v_j = \sum_{\substack{1 \leq j \leq n \\ \epsilon_j = 1}} v_j - \sum_{\substack{1 \leq j \leq n \\ \epsilon_j = -1}} v_j,$$

which implies that

$$(152.8) \quad \left\| \sum_{j=1}^n \epsilon_j v_j \right\| \leq \left\| \sum_{\substack{1 \leq j \leq n \\ \epsilon_j = 1}} v_j \right\| + \left\| \sum_{\substack{1 \leq j \leq n \\ \epsilon_j = -1}} v_j \right\| \leq 2 \|\{v_j\}_{j=1}^\infty\|_{Y(\mathbf{Z}_+, V)}.$$

Thus $\{v_j\}_{j=1}^\infty \in Z(V)$ and

$$(152.9) \quad \|\{v_j\}_{j=1}^\infty\|_{Z(V)} \leq 2 \|\{v_j\}_{j=1}^\infty\|_{Y(\mathbf{Z}_+, V)},$$

which shows that $Y(\mathbf{Z}_+, V) = Z(V)$, and that the corresponding norms are equivalent.

Let $Z_0(V)$ be the closure in $Z(V)$ of the collection of sequences $\{v_j\}_{j=1}^\infty$ of vectors in V with $v_j = 0$ for all but finitely many j . This is the same as

the closure of this set in $Y(\mathbf{Z}_+, V)$, which is also the same as the collection $Y_0(\mathbf{Z}_+, V)$ of sequences $\{v_j\}_{j=1}^\infty$ of elements of V such that $\sum_{j \in \mathbf{Z}_+} v_j$ satisfies the generalized Cauchy criterion. If V is complete, then this is the same as the collection of sequences $\{v_j\}_{j=1}^\infty$ of vectors in V such that $\sum_{j \in \mathbf{Z}_+} v_j$ converges in the generalized sense, as usual. This characterization of $Z_0(V)$ is basically equivalent to the discussion in the previous section.

Suppose that $\{v_j\}_{j=1}^\infty$ is a sequence of vectors in V that is not in $Z(V)$. Thus for each $N \geq 1$ there is an $n \in \mathbf{Z}_+$ and an $\epsilon \in \{1, -1\}^n$ such that

$$(152.10) \quad \left\| \sum_{j=1}^n \epsilon_j v_j \right\| \geq N.$$

Equivalently, for each $l, L \geq 1$ there is an $n \geq l$ and $\epsilon_l, \dots, \epsilon_n \in \{1, -1\}$ such that

$$(152.11) \quad \left\| \sum_{j=l}^n \epsilon_j v_j \right\| \geq L + \sum_{j=1}^{l-1} \|v_j\|.$$

This follows from the previous statement by taking $N = L + 2 \sum_{j=1}^{l-1} \|v_j\|$, and using the triangle inequality to get that

$$(152.12) \quad \left\| \sum_{j=1}^n \epsilon_j v_j \right\| \leq \left\| \sum_{j=l}^n \epsilon_j v_j \right\| + \sum_{j=1}^{l-1} \|v_j\|.$$

Applying (152.11) repeatedly, we get a strictly increasing sequence n_1, n_2, \dots of positive integers and a sequence $\epsilon_1, \epsilon_2, \dots$ with $\epsilon_j \in \{1, -1\}$ for each j such that

$$(152.13) \quad \left\| \sum_{j=1}^{n_1} \epsilon_j v_j \right\| \geq 1$$

and

$$(152.14) \quad \left\| \sum_{j=n_k+1}^{n_{k+1}} \epsilon_j v_j \right\| \geq k + 1 + \sum_{j=1}^{n_k} \|v_j\|$$

for each $k \geq 1$. Using the triangle inequality again, we get that

$$(152.15) \quad \left\| \sum_{j=n_k+1}^{n_{k+1}} \epsilon_j v_j \right\| \leq \left\| \sum_{j=1}^{n_{k+1}} \epsilon_j v_j \right\| + \sum_{j=1}^{n_k} \|v_j\|$$

for each k . Hence

$$(152.16) \quad \left\| \sum_{j=1}^{n_k} \epsilon_j v_j \right\| \geq k$$

for each k , so that the partial sums $\sum_{j=1}^n \epsilon_j v_j$ are not uniformly bounded over $n \in \mathbf{Z}_+$ even for this single sequence $\epsilon = \{\epsilon_j\}_{j=1}^\infty$. If $\{v_j\}_{j=1}^\infty$ is a sequence of vectors in V for which the partial sums $\sum_{j=1}^\infty \epsilon_j v_j$ are uniformly bounded over $n \in \mathbf{Z}_+$ for each sequence $\epsilon = \{\epsilon_j\}_{j=1}^\infty$ of elements of $\{1, -1\}$, then it follows that $\{v_j\}_{j=1}^\infty \in Z(V)$.

153 Bounded coefficients

Let E be a nonempty set, and let V be a real or complex vector space with a norm $\|v\|$. Also let $f \in Y(E, V)$ be given, as in Section 27. If $A \subseteq E$, then we let $\mathbf{1}_A(x)$ be the indicator function associated to A on E , equal to 1 when $x \in A$ and to 0 when $x \in E \setminus A$. Thus

$$(153.1) \quad \sum_{x \in B} \mathbf{1}_A(x) f(x) = \sum_{x \in A \cap B} f(x)$$

for every finite set $B \subseteq E$, which implies that $\mathbf{1}_A f \in Y(E, V)$, and that

$$(153.2) \quad \|\mathbf{1}_A f\|_{Y(E, V)} \leq \|f\|_{Y(E, V)}.$$

Now let a be a real-valued function on E such that $0 \leq a(x) \leq 1$ for every $x \in E$. Let A_1 be the set of $x \in E$ such that $a(x) \geq 1/2$, and put

$$(153.3) \quad a_1(x) = a(x) - \frac{1}{2} \mathbf{1}_{A_1}(x).$$

Thus $0 \leq a_1(x) \leq 1/2$ for every $x \in E$, and we can repeat the process by taking A_2 to be the set of $x \in E$ such that $a_1(x) \geq 1/4$. Continuing in this manner, we get a sequence of subsets A_1, A_2, \dots of E such that

$$(153.4) \quad a(x) = \sum_{j=1}^{\infty} 2^{-j} \mathbf{1}_{A_j}(x)$$

for each $x \in E$. If $f \in Y(E, V)$, as before, then it follows that $a f \in Y(E, V)$ too, and that

$$(153.5) \quad \|a f\|_{Y(E, V)} \leq \|f\|_{Y(E, V)}.$$

If a is a bounded nonnegative real-valued function on E , then we get that $a f \in Y(E, V)$, with

$$(153.6) \quad \|a f\|_{Y(E, V)} \leq \|a\|_{\infty} \|f\|_{Y(E, V)}.$$

If a is any bounded real-valued function on E , then we can apply the previous remarks to the positive and negative parts of a , to get that $a f \in Y(E, V)$ and

$$(153.7) \quad \|a f\|_{Y(E, V)} \leq 2 \|a\|_{\infty} \|f\|_{Y(E, V)}.$$

If V is complex and a is a bounded complex-valued function on E , then we can apply this to the real and imaginary parts of a , to get that $a f \in Y(E, V)$ and

$$(153.8) \quad \|a f\|_{Y(E, V)} \leq 4 \|a\|_{\infty} \|f\|_{Y(E, V)}.$$

In particular, multiplication by a defines a bounded linear operator on $Y(E, V)$ in each case.

Of course, if $f(x) \neq 0$ for only finitely many $x \in E$, then $a f$ has the same property. This implies that $a f \in Y_0(E, V)$ when $f \in Y_0(E, V)$ and a is bounded,

because $Y_0(E, V)$ is the closure in $Y(E, V)$ of the linear subspace of functions on E with finite support. Equivalently, $\sum_{x \in E} a(x) f(x)$ satisfies the generalized Cauchy condition when $\sum_{x \in E} f(x)$ satisfies the generalized Cauchy condition and a is a bounded. If V is complete, then it follows that $\sum_{x \in E} a(x) f(x)$ converges in the generalized sense when $\sum_{x \in E} f(x)$ converges in the generalized sense and a is bounded.

154 Another norm

Let E be a nonempty set, and let V be a real or complex vector space with a norm $\|v\|$. Suppose that $f(x)$ is a V -valued function on E , and consider sums of the form

$$(154.1) \quad \sum_{x \in B} \beta(x) f(x),$$

where $B \subseteq E$ is a nonempty finite set, and β is a function on B with values in $\{1, -1\}$. Of course, this is the same as

$$(154.2) \quad \sum_{x \in B_+} f(x) - \sum_{x \in B_-} f(x),$$

where $B_{\pm} = \{x \in B : \beta(x) = \pm 1\}$. If $Z(E, V)$ is the space of V -valued functions on E for which these sums have bounded norm, then it is easy to see that $Z(E, V)$ is the same as the space $Y(E, V)$ discussed in Section 27. More precisely, $Z(E, V) \subseteq Y(E, V)$ because one can take $\beta(x) = 1$ for each $x \in B$, while $Y(E, V) \subseteq Z(E, V)$ by the triangle inequality. If $f \in Y(E, V) = Z(E, V)$, then put

$$(154.3) \quad \|f\|_{Z(E, V)} = \sup_{B, \beta} \left\| \sum_{x \in B} \beta(x) f(x) \right\|,$$

where the supremum is taken over all nonempty finite sets $B \subseteq E$ and functions $\beta : B \rightarrow \{-1, 1\}$. Note that

$$(154.4) \quad \|f\|_{Y(E, V)} \leq \|f\|_{Z(E, V)} \leq 2 \|f\|_{Y(E, V)},$$

for the same reasons that $Y(E, V) = Z(E, V)$.

If $E = \mathbf{Z}_+$, then the $Z(E, V)$ norm reduces to the $Z(V)$ norm described in Section 152, where we identify V -valued functions on \mathbf{Z}_+ with sequences whose terms are in V . Clearly

$$(154.5) \quad \|f\|_{Z(V)} \leq \|f\|_{Z(\mathbf{Z}_+, V)}$$

for each $f \in Z(V) = Y(\mathbf{Z}_+, V)$, because the $Z(V)$ corresponds to taking B to be of the form $\{1, \dots, n\}$, $n \in \mathbf{Z}_+$, in the previous paragraph. Conversely, if B is any nonempty finite set of positive integers, and $\beta : B \rightarrow \{1, -1\}$, then we can take n to be the maximal element of B , and put $\epsilon_j = \epsilon'_j = \beta(j)$ when $j \in B$, and $\epsilon_j = 1$ and $\epsilon'_j = -1$ when $1 \leq j \leq n$ and $j \notin B$. Thus

$$(154.6) \quad 2 \sum_{j \in B} \beta(j) f(j) = \sum_{j=1}^n \epsilon_j f(j) + \sum_{j=1}^n \epsilon'_j f(j),$$

and hence

$$(154.7) \quad 2 \left\| \sum_{j \in B} \beta(j) f(j) \right\| = \left\| \sum_{j=1}^n \epsilon_j f(j) \right\| + \left\| \sum_{j=1}^n \epsilon'_j f(j) \right\| \leq 2 \|f\|_{Z(V)}.$$

This implies that

$$(154.8) \quad \|f\|_{Z(\mathbf{Z}_+, V)} \leq \|f\|_{Z(V)},$$

by taking the supremum over B, β .

Let E be any nonempty set again, and let A, B be disjoint nonempty finite subsets of E . Also let α, β be functions on A, B , respectively, with values in $\{1, -1\}$. Let γ, γ' be the functions on $C = A \cup B$ defined by $\gamma(x) = \gamma'(x) = \alpha(x)$ when $x \in A$ and $\gamma(x) = -\gamma'(x) = \beta(x)$ when $x \in B$. If $f(x)$ is any V -valued function on E , then

$$(154.9) \quad 2 \sum_{x \in A} \alpha(x) f(x) = \sum_{x \in C} \gamma(x) f(x) + \sum_{x \in C} \gamma'(x) f(x)$$

and

$$(154.10) \quad 2 \sum_{x \in B} \beta(x) f(x) = \sum_{x \in C} \gamma(x) f(x) - \sum_{x \in C} \gamma'(x) f(x).$$

In particular,

$$(154.11) \quad 2 \left\| \sum_{x \in A} \alpha(x) f(x) \right\| \leq \left\| \sum_{x \in C} \gamma(x) f(x) \right\| + \left\| \sum_{x \in C} \gamma'(x) f(x) \right\|,$$

as in the preceding paragraph.

Suppose now that V is uniformly convex, and let $\epsilon > 0$ be given. As in Section 143, there is a $\delta_1 > 0$ such that $\|v - w\| < \epsilon$ whenever $v, w \in V$ satisfy $\|v\|, \|w\| \leq 1$ and $\|(v + w)/2\| > \delta_1$. Equivalently, $\|v - w\| < \epsilon R$ when $\|v\|, \|w\| \leq R$ and $\|(v + w)/2\| > (1 - \delta_1) R$ for any $R > 0$, by dividing by R . Let $f \in Y(E, V)$ with $f \not\equiv 0$ be given, and let us apply this with $R = \|f\|_{Z(E, V)}$. By definition of $\|f\|_{Z(E, V)}$, there is a nonempty finite set $A \subseteq E$ and a function $\alpha : A \rightarrow \{1, -1\}$ such that

$$(154.12) \quad \left\| \sum_{x \in A} \alpha(x) f(x) \right\| > (1 - \delta_1) \|f\|_{Z(E, v)}.$$

Let B be another nonempty finite subset of E that is disjoint from A , and let β be a function on B with values in $\{1, -1\}$. If C, γ , and γ' are as in the previous paragraph and

$$(154.13) \quad v = \sum_{x \in C} \gamma(x) f(x), \quad w = \sum_{x \in C} \gamma'(x) f(x),$$

then $\|v\|, \|w\| \leq \|f\|_{Z(E, V)}$, and

$$(154.14) \quad \left\| \frac{v + w}{2} \right\| = \left\| \sum_{x \in A} \alpha(x) f(x) \right\| > (1 - \delta_1) \|f\|_{Z(E, V)}.$$

Because of uniform convexity, we get that

$$(154.15) \quad 2 \left\| \sum_{x \in B} \beta(x) f(x) \right\| = \|v - w\| < \epsilon \|f\|_{Z(E,V)}.$$

It follows that $\sum_{x \in E} f(x)$ satisfies the generalized Cauchy condition, and hence converges in the generalized sense when V is also complete.

155 Additional properties

Let E be a nonempty set, and let V be a real or complex vector space with a norm $\|v\|$. If $a : E \rightarrow \{1, -1\}$ and $f \in Y(E, V)$, then $af \in Y(E, V)$, and in fact

$$(155.1) \quad \|f\|_{Z(E,V)} = \sup_a \|af\|_{Y(E,V)},$$

where the supremum is taken over all such mappings a . In particular,

$$(155.2) \quad \|bf\|_{Z(E,V)} = \|f\|_{Z(E,V)}$$

for every $f \in Y(E, V)$ and $b : E \rightarrow \{1, -1\}$. If $f \in Y(E, V)$ and b is a bounded real-valued function on E , then $bf \in Y(E, V)$, as in Section 153, and

$$(155.3) \quad \|bf\|_{Z(E,V)} \leq \|b\|_\infty \|f\|_{Z(E,V)}.$$

This follows from the analogous statement for the $Y(E, V)$ norm in Section 153 when b is nonnegative, and otherwise one can express b as the product of a nonnegative function and a function with values in $\{1, -1\}$.

Suppose now that V is a complex vector space, and let \mathbf{T} be the unit circle in the complex plane, consisting of the complex numbers z with $|z| = 1$. If $a : E \rightarrow \mathbf{T}$ and $f \in Y(E, V)$, then $af \in Y(E, V)$ and

$$(155.4) \quad \|af\|_{Y(E,V)} \leq 4 \|f\|_{Y(E,V)},$$

as in Section 153. Put

$$(155.5) \quad \|f\|_{W(E,V)} = \sup_a \|af\|_{Y(E,V)},$$

where the supremum is taken over all mappings $a : E \rightarrow \mathbf{T}$. It is easy to see that this is a norm on $Y(E, V)$, and that

$$(155.6) \quad \|f\|_{Y(E,V)} \leq \|f\|_{W(E,V)} \leq 4 \|f\|_{Y(E,V)}$$

for every $f \in Y(E, V)$. Equivalently,

$$(155.7) \quad \|f\|_{W(E,V)} = \sup_{B, \beta} \left\| \sum_{x \in B} \beta(x) f(x) \right\|,$$

where the supremum is taken over all nonempty finite sets $B \subseteq E$ and functions $\beta : B \rightarrow \mathbf{T}$.

More precisely, one can also check that

$$(155.8) \quad \|f\|_{Z(E,V)} \leq \|f\|_{W(E,V)} \leq 2 \|f\|_{Z(E,V)}$$

for every $f \in Y(E, V)$. The first inequality follows from the definitions and the fact that $1, -1 \in \mathbf{T}$. The second inequality uses the estimate

$$(155.9) \quad \|a f\|_{Z(E,V)} \leq 2 \|a\|_{\infty} \|f\|_{Z(E,V)}$$

for every bounded complex-valued function a on E and $f \in Y(E, V)$. This follows from (155.3) applied to the real and imaginary parts of a .

By construction,

$$(155.10) \quad \|b f\|_{W(E,V)} = \|f\|_{W(E,V)}$$

for every $f \in Y(E, V)$ and $b : E \rightarrow \mathbf{T}$. If b is a bounded complex-valued function on E , then

$$(155.11) \quad \|b f\|_{W(E,V)} \leq \|b\|_{\infty} \|f\|_{W(E,V)}$$

for every $f \in Y(E, V)$. In the case where b is a bounded nonnegative real-valued function on E , this follows from the corresponding statement for the $Y(E, V)$ norm in Section 153. Otherwise, one can express b as the product of a nonnegative real-valued function and a function with values in \mathbf{T} , to get the same conclusion from the previous two cases.

156 Tori

Let \mathbf{T} be the unit circle in the complex plane, as before. It is well known that

$$(156.1) \quad \int_{\mathbf{T}} z |dz| = 0,$$

where $|dz|$ denotes the element of integration with respect to arc length. One way to see this is to compare this integral with a line integral,

$$(156.2) \quad \int_{\mathbf{T}} i z |dz| = \oint_{\mathbf{T}} dz = 0,$$

using the fact that the unit tangent vector to \mathbf{T} at a point $z \in \mathbf{T}$ corresponds to $i z$ with respect to the standard orientation. Alternatively, one can use the change of variables $z \mapsto -z$ to get that

$$(156.3) \quad \int_{\mathbf{T}} z |dz| = - \int_{\mathbf{T}} z |dz|,$$

and hence that the integral is 0, because arc length is not affected by this transformation.

Of course, \mathbf{T} is a compact Hausdorff topological space, and a probability space with respect to arc length measure divided by 2π . As usual, the n -dimensional torus \mathbf{T}^n is the Cartesian product of n copies of \mathbf{T} , consisting of ordered n -tuples $z = (z_1, \dots, z_n)$ with $z_j \in \mathbf{T}$ for $j = 1, \dots, n$. This is also a compact Hausdorff topological space for each n , and a probability space with respect to the corresponding product measure. The coordinate functions z_1, \dots, z_n may be considered as complex-valued independent random variables on \mathbf{T}^n .

Similarly, we can consider the space \mathbf{T}^∞ of sequences $z = \{z_j\}_{j=1}^\infty$ such that $z_j \in \mathbf{T}$ for each j , which is the Cartesian product of a sequence of copies of \mathbf{T} . This is a compact Hausdorff topological space with respect to the product topology, and a probability space with respect to the product measure. The coordinate functions z_1, z_2, \dots form an infinite sequence of independent random variables on this infinite-dimensional torus, as before. Note that the sequences $x = \{x_j\}_{j=1}^\infty$ with $x_j = 1$ or -1 for each j form a closed set in \mathbf{T}^∞ .

Let $(V, \|v\|)$ be a complex Banach space, and let $\{v_j\}_{j=1}^\infty$ be a sequence of elements of V . Consider the V -valued functions

$$(156.4) \quad f_n(z) = \sum_{j=1}^n z_j v_j$$

on \mathbf{T}^∞ for each $n \geq 1$. If $\sum_{j \in \mathbf{Z}_+} v_j$ converges in the generalized sense, then $\{f_n\}_{n=1}^\infty$ converges uniformly on \mathbf{T}^∞ . This is similar to the discussion in Section 151, using also the estimates in Section 153, or the $W(\mathbf{Z}_+, V)$ norm in the previous section, which is basically the same. The converse statements discussed in Section 151 are already applicable in this situation, because $1, -1 \in \mathbf{T}$.

157 Norms and linear functionals

Let E be a nonempty set, and let V be a real or complex vector space with a norm $\|v\|$. If $f \in Y(E, V)$ and λ is a bounded linear functional on V , then $\lambda(f(x))$ is a summable function on E , as in Section 30. Put

$$(157.1) \quad \|f\|_{L(E, V)} = \sup \left\{ \sum_{x \in E} |\lambda(f(x))| : \lambda \in V^*, \|\lambda\|_* \leq 1 \right\}.$$

As in Section 30, this is less than or equal to $2\|f\|_{Y(E, V)}$ in the real case, less than or equal to $4\|f\|_{Y(E, V)}$ in the complex case, and greater than or equal to $\|f\|_{Y(E, V)}$ in both cases. It is easy to see from the definition that $\|f\|_{L(E, V)}$ is a norm on $Y(E, V)$, and that

$$(157.2) \quad \|bf\|_{L(E, V)} \leq \|b\|_\infty \|f\|_{L(E, V)}$$

for every $f \in Y(E, V)$ and bounded real or complex-valued function b on E , as appropriate. If $B \subseteq E$ is a finite set, β is a real or complex-valued function on

B such that $|\beta(x)| = 1$ for each $x \in B$, and $\lambda \in V^*$, then

$$(157.3) \quad \left| \lambda \left(\sum_{x \in B} \beta(x) f(x) \right) \right| = \left| \sum_{x \in B} \beta(x) \lambda(f(x)) \right| \leq \sum_{x \in B} |\lambda(f(x))|,$$

with equality in the last step for suitable choices of β . Using this, one can check that $\|f\|_{L(E,V)}$ is equal to $\|f\|_{Z(E,V)}$ in the real case, and is equal to $\|f\|_{W(E,V)}$ in the complex case.

158 Sums and $c_0(E)$

Let E be a nonempty set, and let $(V, \|v\|)$ be a real or complex Banach space. If $f \in Y(E, V)$ and a is a bounded real or complex-valued function on E , as appropriate, then $a f \in Y(E, V)$ and

$$(158.1) \quad \|a f\|_{Y(E,V)} \leq 2 \|a\|_\infty \|f\|_{Y(E,V)}$$

in the real case, and

$$(158.2) \quad \|a f\|_{Y(E,V)} \leq 4 \|a\|_\infty \|f\|_{Y(E,V)}$$

in the complex case, as in Section 153. If $a \in c_0(E)$, then it follows that $a f$ is in $Y_0(E, V)$, since a can be approximated by functions with finite support in the ℓ^∞ norm. This is the same as saying that $\sum_{x \in E} a(x) f(x)$ satisfies the generalized Cauchy criterion when $a \in c_0(E)$, and hence converges in the generalized sense because V is complete. Thus

$$(158.3) \quad T_f(a) = \sum_{x \in E} a(x) f(x)$$

defines a bounded linear mapping from $c_0(E)$ into V . One can check that the operator norm of T_f is equal to the $Z(E, V)$ norm of f in the real case, and is equal to the $W(E, V)$ norm of f in the complex case. Conversely, if T is a bounded linear mapping from $c_0(E)$ in V , then $T = T_f$ for some $f \in Y(E, V)$. To see this, one can take

$$(158.4) \quad f(x) = T(\delta_x),$$

where δ_x is the function on E defined by $\delta_x(x) = 1$ and $\delta_x(y) = 0$ when $x \neq y$. If a is a real or complex-valued function on E with finite support, then a is a linear combination of finitely many δ_x 's, and so $T(a)$ is given by the same expression as $T_f(a)$, because of linearity. Using this and the boundedness of T , one can show that $f \in Y(E, V)$, and more precisely that the $Z(E, V)$ norm of f is less than or equal to the operator norm of T in the real case, and that the $W(E, V)$ norm of f is less than or equal to the operator norm of T in the complex case. This implies that $T(a) = T_f(a)$ for every $a \in c_0(E)$, because T, T_f are bounded linear operators which agree on the dense linear subspace of $c_0(E)$ consisting of functions a with finite support.

159 Integrability

Let (X, \mathcal{A}, μ) be a measure space, and let $(V, \|v\|)$ be a real or complex Banach space. As in Section 120, it is easy to deal with integration of functions with values in a finite-dimensional subspace of V . Suppose that $\{f_j\}_{j=1}^\infty$ is a sequence of V -valued functions on X such that each f_j takes values in a finite-dimensional subspace of V , each f_j is integrable in the sense of Section 120, and

$$(159.1) \quad \lim_{j, l \rightarrow \infty} \int_X \|f_j - f_l\| d\mu = 0.$$

This implies in particular that the sequence of integrals

$$(159.2) \quad \int_X f_j d\mu$$

is a Cauchy sequence in V , and hence converges in V , by completeness.

A sufficient condition for this type of convergence to hold is that

$$(159.3) \quad \sum_{j=1}^{\infty} \int_X \|f_j - f_{j+1}\| d\mu < \infty.$$

This is the same as

$$(159.4) \quad \int_X \sum_{j=1}^{\infty} \|f_j - f_{j+1}\| d\mu < \infty,$$

which implies that

$$(159.5) \quad \sum_{j=1}^{\infty} \|f_j(x) - f_{j+1}(x)\| < \infty$$

for almost every $x \in X$. It follows that

$$(159.6) \quad \sum_{j=1}^{\infty} (f_j(x) - f_{j+1}(x))$$

converges in V for almost every $x \in X$, by completeness again. Put

$$(159.7) \quad f(x) = \lim_{j \rightarrow \infty} f_j(x),$$

which exists for almost every $x \in X$ by the convergence of the previous sum. Of course, any sequence of V -valued functions as in the preceding paragraph has a subsequence that satisfies this summability condition, and hence converges almost everywhere.

Under these conditions, put

$$(159.8) \quad \int_X f d\mu = \lim_{j \rightarrow \infty} \int_X f_j d\mu.$$

This is basically the definition of the Bochner integral. Note that $\{\|f_j(x)\|\}_{j=1}^\infty$ converges in $L^1(X)$ to $\|f(x)\|$, which implies that

$$(159.9) \quad \left\| \int_X f \, d\mu \right\| \leq \int_X \|f\| \, d\mu.$$

Similarly, if λ is a bounded linear functional on V , then $\lambda(f_j(x))$ converges in $L^1(X)$ to $\lambda(f(x))$, and hence

$$(159.10) \quad \lambda\left(\int_X f \, d\mu\right) = \int_X \lambda \circ f \, d\mu.$$

This shows that the integral of f does not depend on the particular sequence of approximations.

Remember that a function on X with values in a topological space is said to be measurable if the inverse image of every open set in the range is measurable. Thus the composition of a measurable function with a continuous mapping to another topological space is also measurable. If $f : X \rightarrow V$ is measurable with respect to the topology on V associated to the norm, then it follows that $\|f(x)\|$ is measurable too. If in addition $\|f(x)\|$ is integrable and V is separable, then f can be approximated by integrable functions with values in finite-dimensional subspaces of V , as before. To see this, one can start by using the integrability of $\|f(x)\|$ to approximate f by bounded measurable V -valued functions that are equal to 0 on the complements of suitable subsets of finite measure. One can then use the separability of V to approximate these functions by V -valued simple functions. The same argument would work if f takes values in a separable subspace of V almost everywhere on X .

160 Bounded measures

Let X be a set, let \mathcal{A} be an algebra of subsets of X , and let V be a real or complex vector space. A V -valued function μ on \mathcal{A} is said to be a finitely-additive V -valued measure on (X, \mathcal{A}) if

$$(160.1) \quad \mu(A \cup B) = \mu(A) + \mu(B)$$

for every $A, B \in \mathcal{A}$ with $A \cap B = \emptyset$. If $A_1, \dots, A_n \in \mathcal{A}$ and $t_1, \dots, t_n \in \mathbf{R}$ or \mathbf{C} , as appropriate, then

$$(160.2) \quad f(x) = \sum_{j=1}^n t_j \mathbf{1}_{A_j}(x)$$

is a measurable simple function on X , and we put

$$(160.3) \quad \int_X f \, d\mu = \sum_{j=1}^n t_j \mu(A_j).$$

It is easy to see that this does not depend on the particular representation of f as a linear combination of indicator functions, and that it defines a linear

mapping from the vector space of measurable simple functions on X into V . If λ is a linear functional on V , then $\mu_\lambda(A) = \lambda(\mu(A))$ is a finitely-additive real or complex measure on (X, \mathcal{A}) , as appropriate, and

$$(160.4) \quad \lambda\left(\int_X f d\mu\right) = \int_X f d\mu_\lambda.$$

Suppose now that V is equipped with a norm $\|v\|$, and that μ is bounded, so that

$$(160.5) \quad C(\mu) = \sup\{\|\mu(A)\| : A \in \mathcal{A}\} < \infty.$$

If A_1, \dots, A_n are finitely many pairwise-disjoint measurable subsets of X and $E_n = \{1, \dots, n\}$, then $\mu(A_j)$ may be considered as a V -valued function on E_n whose $Y(E_n, V)$ norm is less than or equal to $C(\mu)$, because of the finite additivity of μ . As in Section 153, it follows that

$$(160.6) \quad \left\| \int_X f d\mu \right\| \leq k C(\mu) \sup_{x \in X} |f(x)|$$

for every measurable simple function f on X , where $k = 1$ when f is real-valued and nonnegative, $k = 2$ when f is real-valued, and $k = 4$ when f is complex-valued. If \mathcal{A} is a σ -algebra and V is complete, then the integral can be extended to bounded measurable real or complex-valued functions f on X , as appropriate, because simple functions are dense in the space of bounded measurable functions with respect to the supremum norm. If λ is a bounded linear functional on V , then μ_λ is a bounded finitely-additive real or complex measure on (X, \mathcal{A}) , with

$$(160.7) \quad C(\mu_\lambda) \leq \|\lambda\|_* C(\mu),$$

and we get the same relationship with the integral of a bounded measurable function as for simple functions.

In particular, this works when \mathcal{A} is a σ -algebra and μ is countably additive, in the sense that

$$(160.8) \quad \sum_{j=1}^{\infty} \mu(A_j) = \mu\left(\bigcup_{j=1}^{\infty} A_j\right)$$

for every sequence A_1, A_2, \dots of pairwise-disjoint measurable subsets of X , as in Section 37. More precisely, convergence of the series on the left in V is part of the hypothesis, and we have seen that this implies that μ is bounded. In this case, μ_λ is a countably-additive real or complex measure on (X, \mathcal{A}) for each $\lambda \in V^*$, as before. If μ has the additional property that $\sum_{j=1}^{\infty} \|\mu(A_j)\|$ converges for every sequence A_1, A_2, \dots of pairwise-disjoint measurable subsets of X , then one can integrate any $f \in L^1(X, \|\mu\|)$, as in Section 133. If instead V is a Hilbert space and $\mu(A)$ is orthogonal to $\mu(B)$ when A, B are disjoint measurable subsets of X , then the integral can be defined on a suitable L^2 space, as in Section 134.

As another situation like this, suppose that $V = W^*$ for some Banach space W , \mathcal{A} is a σ -algebra, and μ is countably additive with convergence in the weak*

topology on V . This implies that $\mu_w(A) = \mu(A)(w)$ is a countably-additive real or complex measure on (X, \mathcal{A}) for each $w \in W$. This is the same as the measure μ_λ defined before, where λ is the bounded linear functional on V corresponding to evaluation at w . Using the uniform boundedness principle, one can show that μ is bounded, as in Section 37. Under these conditions, the integral of a bounded measurable function f on X can be defined more directly as a bounded linear functional on W by

$$(160.9) \quad \left(\int_X f \, d\mu \right)(w) = \int_X f \, d\mu_w,$$

which is also satisfied by the previous definition.

161 Weak* measurability

Let W be a real or complex vector space with a norm $\|w\|_W$ which is separable, and let w_1, w_2, \dots be a sequence of elements of W such that $\|w_j\|_W = 1$ for each j and the set of w_j 's is dense in the unit sphere in W . This uses the fact that a subset of a separable metric space is also separable. Note that

$$(161.1) \quad \|\lambda\|_{W^*} = \sup_{j \geq 1} |\lambda(w_j)|$$

for every bounded linear functional λ on W . Also let (X, \mathcal{A}) be a measurable space, and let f be a function on X with values in W^* . If $f(x)(w)$ is measurable as a real or complex-valued function on X for every $w \in W$, then it follows that $\|f(x)\|_V$ is measurable on X as well.

Let us say that $f : X \rightarrow W^*$ is weak* measurable if f is measurable with respect to the weak* topology on W^* . This automatically implies that $f(x)(w)$ is measurable for each $w \in W$, since evaluation at w is a continuous function on W^* . Conversely, f is weak* measurable when $f(x)(w)$ is measurable for every $w \in W$ and W is separable. To see this, one may as well suppose that f is bounded, because one can use the measurability of $\|f(x)\|_{W^*}$ to express X as the union of a sequence of measurable sets on which f is bounded. If B is a ball in W^* , then the topology on B induced by the weak* topology on W^* is metrizable, because W is separable, as in Section 33. If B is a closed ball in W^* , then B is also compact in the weak* topology, by the Banach–Alaoglu theorem. Thus B is compact and metrizable with respect to the topology induced by the weak* topology, and hence is separable with respect to this topology. This implies that relatively open subsets of B in the weak* topology can be given in terms of countable unions of basic open sets, which permits the weak* measurability of f to be obtained from the measurability of $f(x)(w)$ for each $w \in W$.

Of course, f is weak* measurable if f is measurable with respect to the topology on W^* associated to the dual norm, because every open set in W^* with respect to the weak* topology is also open in the norm topology. Conversely, if f is weak* measurable and W^* is separable, then f is measurable with respect to the norm topology on W^* . Indeed, separability of W^* implies that each open

set $U \subseteq W^*$ in the norm topology is a countable union of closed balls. If B is a closed ball in W^* , then B is a closed set in W^* in the weak* topology by the definition of the dual norm, and so $f^{-1}(B)$ is measurable in X by weak* measurability. It follows that $f^{-1}(U)$ is the union of countably many measurable subsets of X , and hence is measurable.

Similarly, if V is a real or complex vector space with a norm $\|v\|_V$, then we say that $f : X \rightarrow V$ is weakly measurable if f is measurable with respect to the weak topology on V . If f is measurable with respect to the topology on V associated to the norm, then f is weakly measurable, because every open set in V with respect to the weak topology is also an open set in the norm topology. Conversely, if f is weakly measurable and V is separable, then f is measurable with respect to norm topology on V . As before, separability of V implies that every open set $U \subseteq V$ in the norm topology is the countable union of closed balls. In this case, the fact that a closed ball B in V is also closed in the weak topology uses the Hahn–Banach theorem. If f is weakly measurable, then it follows that $f^{-1}(B)$ is a measurable set in X for each closed ball B in V , and hence that $f^{-1}(U)$ is measurable in X for each open set $U \subseteq V$ in the norm topology. If V^* is separable, then one can argue as before that $\|f(x)\|$ is measurable on X when $\lambda(f(x))$ is measurable for each $\lambda \in V^*$. The same argument shows that $\|f(x) - v\|$ is measurable on X for every $v \in V$ under these conditions, so that $f^{-1}(B)$ is measurable in X for each ball B in V . One can then use separability of V again to get that f is measurable with respect to the norm topology on V .

162 Weak* measures

Let (X, \mathcal{A}) be a measurable space, and let $(W, \|w\|_W)$ be a real or complex Banach space. Let us say that a function μ on \mathcal{A} with values in the dual W^* of W is a *weak* measure* if

$$(162.1) \quad \sum_{j=1}^{\infty} \mu(A_j) = \mu\left(\bigcup_{j=1}^{\infty} A_j\right)$$

for every sequence A_1, A_2, \dots of pairwise-disjoint measurable subsets of X , where the series is supposed to converge in the weak* topology on W^* . This is equivalent to asking that μ be finitely additive, and that

$$(162.2) \quad \lim_{j \rightarrow \infty} \mu(B_j) = \mu\left(\bigcup_{j=1}^{\infty} B_j\right)$$

in the weak* topology for every increasing sequence B_1, B_2, \dots of measurable subsets of X . This is also equivalent to the condition that μ be finitely additive and satisfy

$$(162.3) \quad \lim_{j \rightarrow \infty} \mu(C_j) = \mu\left(\bigcap_{j=1}^{\infty} C_j\right)$$

in the weak* topology for every decreasing sequence C_1, C_2, \dots of measurable subsets of X . This is also the same as saying that

$$(162.4) \quad \mu_w(A) = \mu(A)(w)$$

is a countably-additive real or complex measure on (X, \mathcal{A}) , as appropriate, for every $w \in W$.

Remember that convergent sequences in W^* in the weak* topology are bounded with respect to the dual norm when W is complete, by the theorem of Banach and Steinhaus. If μ is a weak* measure on (X, \mathcal{A}) with values in W^* , then there is a $C \geq 0$ such that

$$(162.5) \quad \|\mu(A)\|_{W^*} \leq C$$

for every $A \in \mathcal{A}$, by the same arguments as in Section 37. Equivalently,

$$(162.6) \quad |\mu_w(A)| \leq C \|w\|_W$$

for every $w \in W$ and $A \in \mathcal{A}$, which implies that

$$(162.7) \quad |\mu_w|(X) \leq k C \|w\|_W$$

for every $w \in W$, where $k = 2$ in the real case and $k = 4$ in the complex case. Thus $w \mapsto \mu_w$ defines a bounded linear mapping from W into the space of real or complex measures on (X, \mathcal{A}) , as appropriate, equipped with the norm associated to the total variation. Conversely, a bounded linear mapping from W into the space of real or complex measures on (X, \mathcal{A}) determines a weak* measure on (X, \mathcal{A}) with values in W^* in this way.

If ν is a nonnegative measure on (X, \mathcal{A}) and $f \in L^1(X, \nu)$, then

$$(162.8) \quad \nu_f(A) = \int_A f d\nu$$

defines a real or complex measure on (X, \mathcal{A}) , as appropriate. Thus a bounded linear mapping from W into $L^1(X, \nu)$ determines a weak* measure μ on (X, \mathcal{A}) with values in W^* , as in the previous paragraph. In this case, μ is absolutely continuous with respect to ν , in the sense that $\mu(A) = 0$ for every measurable set $A \subseteq X$ with $\nu(A) = 0$, because ν_f is absolutely continuous with respect to ν for every $f \in L^1(X, \nu)$. Of course, any weak* measure μ on (X, \mathcal{A}) with values in W^* is absolutely continuous with respect to ν in this sense if and only if μ_w is absolutely continuous with respect to ν for each $w \in W$. If ν is σ -finite, then the Radon–Nikodym theorem implies that every weak* measure μ on (X, \mathcal{A}) that is absolutely continuous with respect to ν corresponds to a bounded linear mapping from W into $L^1(X, \nu)$.

If E is a nonempty set, then $Y(E, W^*)$ can be identified with the space of bounded linear mappings from W into $\ell^1(E)$. This is basically another way of looking at the discussion in Section 32. We can also think of $\ell^1(E)$ as being the L^1 space associated to counting measure on E , so that elements of $\ell^1(E)$

determine real or complex measures on E as in the preceding paragraph. More precisely, these are measures defined on arbitrary subsets of E . It follows that elements of $Y(E, W^*)$ determine bounded linear mappings from W into real or complex measures on E , as appropriate, and hence weak* measures on E with values in W^* .

163 Weak* integrability

Let (X, \mathcal{A}, ν) be a measure space, let W be a real or complex vector space with a norm $\|w\|_W$. Also let f be a W^* -valued function on X such that $f(x)(w)$ is measurable on X for each $w \in W$. If W is separable, then it follows that $\|f(x)\|_{W^*}$ is measurable on X , as in Section 161. Alternatively, if $f : X \rightarrow W^*$ is measurable with respect to the weak* topology on W^* , then we get that $f(x)(w)$ is measurable for each $w \in W$ and that $\|f(x)\|_{W^*}$ is measurable. The latter uses the fact that closed balls in W^* are closed sets in the weak* topology, by definition of the dual norm.

At any rate, if $\|f(x)\|_{W^*}$ is integrable with respect to ν , then $f(x)(w)$ is also integrable with respect to ν for each $w \in W$, and

$$(163.1) \quad \int_X |f(x)(w)| d\nu(x) \leq \|w\|_W \int_X \|f(x)\|_{W^*} d\nu(x)$$

for every $w \in W$. In particular, $w \mapsto f(x)(w)$ is a bounded linear mapping from W into $L^1(X, \nu)$, which leads to a weak* measure μ on (X, \mathcal{A}) with values in W^* , as in the previous section. More precisely,

$$(163.2) \quad \mu_w(A) = \mu(A)(w) = \int_A f(x)(w) d\nu(x)$$

for every measurable set $A \subseteq X$ and $w \in W$, which implies that

$$(163.3) \quad \|\mu(A)\|_{W^*} \leq \int_A \|f(x)\|_{W^*} d\nu.$$

If A_1, A_2, \dots is a sequence of pairwise-disjoint measurable subsets of X , then it is easy to see that $\sum_{j=1}^{\infty} \mu(A_j)$ converges absolutely with respect to the dual norm on W^* , and that the sum is equal to $\mu(\bigcup_{j=1}^{\infty} A_j)$.

Let $\|\mu\|(A)$ be the total variation measure associated to μ as in Section 37. Thus $\|\mu\|(A) = p^*(A)$ corresponds to $p(A) = \|\mu(A)\|_{W^*}$ as in Section 35. In this case,

$$(163.4) \quad \|\mu\|(A) \leq \int_A \|f(x)\|_{W^*} d\nu(x)$$

for each $A \in \mathcal{A}$, because of (163.3). Of course,

$$(163.5) \quad |\mu_w(A)| \leq \|\mu(A)\|_{W^*} \|w\|_W \leq \|\mu\|(A) \|w\|_W$$

for every $A \in \mathcal{A}$ and $w \in W$, which implies that

$$(163.6) \quad |\mu_w|(A) \leq \|\mu\|(A) \|w\|_W,$$

where $|\mu_w|$ is the total variation measure associated to μ_w . Hence

$$(163.7) \quad \int_A |f(x)(w)| d\nu(x) \leq \|\mu\|(A) \|w\|$$

for every $A \in \mathcal{A}$ and $w \in W$.

Suppose that W is separable, and let w_1, w_2, \dots be a sequence of elements of W such that $\|w_j\|_W = 1$ for each j and the set of w_j 's is dense in the unit sphere in W . If

$$(163.8) \quad \phi_n(x) = \max_{1 \leq j \leq n} |f_j(x)(w_j)|,$$

then $\phi_n(x)$ is measurable on X for each n ,

$$(163.9) \quad \phi_n(x) \leq \phi_{n+1}(x) \leq \|f(x)\|_{W^*},$$

for each $x \in X$ and $n \geq 1$, and

$$(163.10) \quad \lim_{n \rightarrow \infty} \phi_n(x) = \sup_{n \geq 1} \phi_n(x) = \|f(x)\|_{W^*}$$

for each $x \in X$. Let $A \subseteq X$ be a measurable set, and let A_1, \dots, A_n be pairwise-disjoint measurable subsets of X such that $\bigcup_{j=1}^n A_j = A$. Observe that

$$(163.11) \quad \sum_{j=1}^n \int_{A_j} |f(x)(w_j)| d\nu(x) \leq \sum_{j=1}^n \|\mu\|(A_j) = \|\mu\|(A).$$

This implies that

$$(163.12) \quad \int_A \phi_n(x) d\nu(x) \leq \|\mu\|(A)$$

for each n . Using the monotone convergence theorem, we get that

$$(163.13) \quad \int_A \|f(x)\|_{W^*} d\nu(x) \leq \|\mu\|(A).$$

It follows that

$$(163.14) \quad \|\mu\|(A) = \int_A \|f(x)\|_{W^*} d\nu(x)$$

for every $A \in \mathcal{A}$ when W is separable.

References

- [1] M. Adams and V. Guillemin, *Measure Theory and Probability*, Birkhäuser, 1996.
- [2] F. Albiac and N. Kalton, *Topics in Banach Space Theory*, Springer-Verlag, 2006.
- [3] L. Ambrosio and P. Tilli, *Topics on Analysis in Metric Spaces*, Oxford University Press, 2004.

- [4] R. Ash, *Probability and Measure Theory*, 2nd edition, with contributions by C. Doléans-Dade, Academic Press, 2000.
- [5] K. Athreya and S. Lahiri, *Measure Theory and Probability Theory*, Springer-Verlag, 2006.
- [6] S. Athreya and V. Sunder, *Measure & Probability*, Universities Press and CRC Press, 2008.
- [7] R. Beals, *Advanced Mathematical Analysis*, Springer-Verlag, 1973.
- [8] R. Beals, *Analysis: An Introduction*, Cambridge University Press, 2004.
- [9] A. Beardon, *Limits: A New Approach to Real Analysis*, Springer-Verlag, 1997.
- [10] B. Beauzamy, *Introduction to Banach Spaces and their Geometry*, 2nd edition, North-Holland, 1985.
- [11] Y. Benyamini and Lindenstrauss, *Geometric Nonlinear Functional Analysis*, American Mathematical Society, 2000.
- [12] R. Bhatia, *Notes on Functional Analysis*, Hindustan Book Agency, 2009.
- [13] R. Bhattacharya and E. Waymire, *A Basic Course in Probability Theory*, Springer-Verlag, 2007.
- [14] P. Billingsley, *Ergodic Theory and Information*, Krieger, 1978.
- [15] P. Billingsley, *Probability and Measure*, 3rd edition, Wiley, 1995.
- [16] A. Bobrowski, *Functional Analysis for Probability and Stochastic Processes*, Cambridge University Press, 2005.
- [17] V. Borkar, *Probability Theory*, Springer-Verlag, 1995.
- [18] J. Bourgain, *New Classes of L^p Spaces*, Lecture Notes in Mathematics **889**, Springer-Verlag, 1981.
- [19] R. Bourgin, *Geometric Aspects of Convex Sets with the Radon–Nikodym Property*, Lecture Notes in Mathematics **993**, Springer-Verlag, 1983.
- [20] C. Burrill, *Measure, Integration, and Probability*, McGraw-Hill, 1972.
- [21] M. Capiński and E. Kopp, *Measure, Integral, and Probability*, 2nd edition, Springer-Verlag, 2004.
- [22] N. Carothers, *Real Analysis*, Cambridge University Press, 2000.
- [23] N. Carothers, *A Short Course on Banach Space Theory*, Cambridge University Press, 2005.

- [24] K.-L. Chung, *A Course in Probability Theory*, 3rd edition, Academic Press, 2001.
- [25] K.-L. Chung and F. AitSahlia, *Elementary Probability Theory*, 4th edition, Springer-Verlag, 2003.
- [26] D. Cohn, *Measure Theory*, Birkhäuser, 1993.
- [27] R. Coifman and G. Weiss, *Analyse Harmonique Non-Commutative sur Certains Espaces Homogènes*, Lecture Notes in Mathematics **242**, Springer-Verlag, 1971.
- [28] R. Coifman and G. Weiss, *Transference Methods in Analysis*, American Mathematical Society, 1976.
- [29] R. Coifman and G. Weiss, *Extensions of Hardy spaces and their use in analysis*, Bulletin of the American Mathematical Society **83** (1977), 569–645.
- [30] J. Conway, *A Course in Functional Analysis*, 2nd edition, Springer-Verlag, 1990.
- [31] B. Craven, *Lebesgue Measure & Integral*, Pitman, 1982.
- [32] J. Diestel, *Geometry of Banach Spaces — Selected Topics*, Lecture Notes in Mathematics **485**, Springer-Verlag, 1975.
- [33] J. Diestel, *Sequences and Series in Banach Spaces*, Springer-Verlag, 1984.
- [34] J. Diestel and J. Uhl, Jr., *Vector Measures*, with a foreword by B. Pettis, American Mathematical Society, 1977.
- [35] J. Doob, *Measure Theory*, Springer-Verlag, 1994.
- [36] J. Doob, *Classical Potential Theory and its Probabilistic Counterpart*, Springer-Verlag, 2001.
- [37] R. Dudley, *Real Analysis and Probability*, Cambridge University Press, 2002.
- [38] J. Duoandikoetxea, *Fourier Analysis*, translated and revised from the 1995 Spanish original by D. Cruz-Urbe, American Mathematical Society, 2001.
- [39] P. Duren, *Theory of H^p Spaces*, Academic Press, 1970.
- [40] G. Edgar, *Integral, Probability, and Fractal Measures*, Springer-Verlag, 1998.
- [41] G. Edgar, *Measure, Topology, and Fractal Geometry*, 2nd edition, Springer-Verlag, 2008.

- [42] Y. Eidelman, V. Milman, and A. Tsolomitis, *Functional Analysis: An Introduction*, American Mathematical Society, 2004.
- [43] L. Evans and R. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press, 1992.
- [44] K. Falconer, *The Geometry of Fractal Sets*, Cambridge University Press, 1986.
- [45] H. Federer, *Geometric Measure Theory*, Springer-Verlag, 1969.
- [46] G. Folland, *Real Analysis*, 2nd edition, Wiley, 1999.
- [47] G. Folland, *A Guide to Advanced Real Analysis*, Mathematical Association of America, 2009.
- [48] J. Galambos, *Advanced Probability Theory*, 2nd edition, Dekker, 1995.
- [49] T. Gamelin and R. Greene, *Introduction to Topology*, 2nd edition, Dover, 1999.
- [50] J. García-Cuerva and J. Rubio de Francia, *Weighted Norm Inequalities and Related Topics*, North-Holland, 1985.
- [51] J. Garnett, *Bounded Analytic Functions*, Springer-Verlag, 2007.
- [52] E. Giusti, *Minimal Surfaces and Functions of Bounded Variation*, Birkhäuser, 1984.
- [53] R. Goldberg, *Methods of Real Analysis*, 2nd edition, Wiley, 1976.
- [54] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, 3rd edition, Oxford University Press, 2001.
- [55] G. Grimmett and D. Welsh, *Probability: An Introduction*, Oxford University Press, 1986.
- [56] A. Gut, *Probability: A Graduate Course*, Springer-Verlag, 2005.
- [57] A. Gut, *An Intermediate Course in Probability*, 2nd edition, Springer-Verlag, 2009.
- [58] M. de Guzman, *Differentiation of Integrals in \mathbf{R}^n* , with appendices by A. Córdoba, R. Fefferman, and R. Moriyón, *Lecture Notes in Mathematics* **481**, Springer-Verlag, 1975.
- [59] M. de Guzman, *Real Variable Methods in Fourier Analysis*, North-Holland, 1981.
- [60] P. Halmos, *Measure Theory*, van Nostrand, 1950.
- [61] P. Halmos, *Lectures on Ergodic Theory*, Chelsea, 1960.

- [62] P. Halmos, *A Hilbert Space Problem Book*, 2nd edition, Springer-Verlag, 1982.
- [63] P. Halmos, *Introduction to Hilbert Space and the Theory of Spectral Multiplicity*, AMS Chelsea, 1998.
- [64] V. Hansen, *Fundamental Concepts in Modern Analysis*, World Scientific, 1999.
- [65] V. Hansen, *Functional Analysis: Entering Hilbert Space*, World Scientific, 2006.
- [66] J. Heinonen, *Lectures on Analysis on Metric Spaces*, Springer-Verlag, 2001.
- [67] E. Hernández and G. Weiss, *A First Course on Wavelets*, with a foreword by Y. Meyer, CRC Press, 1996.
- [68] E. Hewitt and K. Stromberg, *Real and Abstract Analysis*, Springer-Verlag, 1975.
- [69] J. Hocking and G. Young, *Topology*, 2nd edition, Dover, 1988.
- [70] K. Hoffman, *Banach Spaces of Analytic Functions*, Dover, 1988.
- [71] K. Itô, *Introduction to Probability Theory*, translated from the Japanese by the author, Cambridge University Press, 1984.
- [72] J. Jacod and P. Protter, *Probability Essentials*, 2nd edition, Springer-Verlag, 2003.
- [73] F. Jones, *Lebesgue Integration on Euclidean Spaces*, Jones and Bartlett, 1993.
- [74] J.-L. Journé, *Calderón–Zygmund Operators, Pseudodifferential Operators, and the Cauchy Integral of Calderón*, Lecture Notes in Mathematics **994**, Springer-Verlag, 1983.
- [75] M. Kadets and V. Kadets, *Series in Banach Spaces*, translated from the Russian by A. Iacob, Birkhäuser, 1997.
- [76] J.-P. Kahane, *Séries de Fourier Absolument Convergentes*, Springer-Verlag, 1970.
- [77] J.-P. Kahane, *Some Random Series of Functions*, 2nd edition, Cambridge University Press, 1985.
- [78] O. Kallenberg, *Foundations of Modern Probability*, 2nd edition, Springer-Verlag, 2002.
- [79] N. Kalton, N. Peck, and J. Roberts, *An F -Space Sampler*, Cambridge University Press, 1984.

- [80] S. Kantorovitz, *Introduction to Modern Analysis*, Oxford University Press, 2003.
- [81] I. Kaplansky, *Set Theory and Metric Spaces*, 2nd edition, Chelsea, 1977.
- [82] A. Karr, *Probability*, Springer-Verlag, 1993.
- [83] Y. Katznelson, *An Introduction to Harmonic Analysis*, 3rd edition, Cambridge University Press, 2004.
- [84] J. Kelley, *General Topology*, Springer-Verlag, 1975.
- [85] J. Kelley, I. Namioka, et al., *Linear Topological Spaces*, Springer-Verlag, 1976.
- [86] J. Kelley and T. Srinivasan, *Measure and Integral*, Springer-Verlag, 1988.
- [87] S. Kesavan, *Functional Analysis*, Hindustan Book Agency, 2009.
- [88] J. Kingman and S. Taylor, *Introduction to Measure and Probability*, Cambridge University Press, 1966.
- [89] A. Klenke, *Probability Theory: A Comprehensive Course*, translated from the 2006 German original, Springer-Verlag, 2008.
- [90] A. Knapp, *Basic Real Analysis*, Birkhäuser, 2005.
- [91] A. Knapp, *Advanced Real Analysis*, Birkhäuser, 2005.
- [92] P. Koosis, *Introduction to H_p Spaces*, 2nd edition, with two appendices by V. Havin, Cambridge University Press, 1998.
- [93] L. Koralov and Y. Sinai, *Theory of Probability and Random Processes*, 2nd edition, Springer-Verlag, 2007.
- [94] S. Krantz, *A Panorama of Harmonic Analysis*, Mathematical Association of America, 1999.
- [95] S. Krantz, *Real Analysis and Foundations*, 2nd edition, Chapman & Hall / CRC, 2005.
- [96] S. Krantz, *A Guide to Real Variables*, Mathematical Association of America, 2009.
- [97] S. Krantz, *A Guide to Topology*, Mathematical Association of America, 2009.
- [98] S. Krantz and H. Parks, *The Geometry of Domains in Space*, Birkhäuser, 1999.
- [99] J. Lamperti, *Probability: A Survey of the Mathematical Theory*, 2nd edition, Wiley, 1996.

- [100] S. Lang, *Real and Functional Analysis*, 3rd edition, Springer-Verlag, 1993.
- [101] S. Lang, *Undergraduate Analysis*, 2nd edition, Springer-Verlag, 1997.
- [102] P. Lax, *Functional Analysis*, Wiley, 2002.
- [103] J. Lindenstrauss and L. Tzafriri, *Classical Banach Spaces*, Lecture Notes in Mathematics **338**, Springer-Verlag, 1973.
- [104] J. Lindenstrauss and L. Tzafriri, *Classical Banach Spaces, I: Sequence Spaces*, Springer-Verlag, 1977.
- [105] J. Lindenstrauss and L. Tzafriri, *Classical Banach Spaces, II: Function Spaces*, Springer-Verlag, 1979.
- [106] B. MacCluer, *Elementary Functional Analysis*, Springer-Verlag, 2009.
- [107] R. Mañé, *Ergodic Theory and Differentiable Dynamics*, translated from the Portuguese by S. Levy, Springer-Verlag, 1987.
- [108] M. Marcus and G. Pisier, *Random Fourier Series with Applications to Harmonic Analysis*, Princeton University Press, 1981.
- [109] P. Mattila, *Geometry of Sets and Measures on Euclidean Spaces*, Cambridge University Press, 1995.
- [110] R. Meester, *A Natural Introduction to Probability Theory*, 2nd edition, Birkhäuser, 2008.
- [111] R. Megginson, *An Introduction to Banach Space Theory*, Springer-Verlag, 1998.
- [112] B. Mendelson, *Introduction to Topology*, 3rd edition, Dover, 1990.
- [113] J. Mikusiński, *The Bochner Integral*, Birkhäuser, 1978.
- [114] T. Morrison, *Functional Analysis: An Introduction to Banach Space Theory*, Wiley, 2001.
- [115] L. Nachbin, *Introduction to Functional Analysis: Banach Spaces and Differential Calculus*, translated from the Portuguese by R. Aron, Dekker, 1981.
- [116] A. Papadopoulos, *Metric Spaces, Convexity and Nonpositive Curvature*, European Mathematical Society, 2005.
- [117] K. Parthasarathy, *Probability Measures on Metric Spaces*, AMS Chelsea, 2005.
- [118] K. Parthasarathy, *Introduction to Probability and Measure*, Hindustan Book Agency, 2005.

- [119] K. Petersen, *Ergodic Theory*, Cambridge University Press, 1989.
- [120] A. Pietsch, *History of Banach Spaces and Linear Operators*, Birkhäuser, 2007.
- [121] G. Pisier, *The Volume of Convex Bodies and Banach Space Geometry*, Cambridge University Press, 1989.
- [122] D. Pollard, *A User's Guide to Measure Theoretic Probability*, Cambridge University Press, 2002.
- [123] D. Promislow, *A First Course in Functional Analysis*, Wiley, 2008.
- [124] S. Resnick, *A Probability Path*, Birkhäuser, 1999.
- [125] F. Riesz and B. Sz. Nagy, *Functional Analysis*, translated from the second French edition by L. Boron, Dover, 1990.
- [126] J. Rosenthal, *A First Look at Rigorous Probability Theory*, 2nd edition, World Scientific, 2006.
- [127] G. Roussas, *An Introduction to Measure-Theoretic Probability*, Elsevier / Academic Press, 2005.
- [128] H. Royden, *Real Analysis*, 3rd edition, Macmillan, 1988.
- [129] W. Rudin, *Principles of Mathematical Analysis*, 3rd edition, McGraw-Hill, 1976.
- [130] W. Rudin, *Real and Complex Analysis*, 3rd edition, McGraw-Hill, 1987.
- [131] W. Rudin, *Fourier Analysis on Groups*, Wiley, 1990.
- [132] W. Rudin, *Functional Analysis*, 2nd edition, McGraw-Hill, 1991.
- [133] B. Rynne and M. Youngson, *Linear Functional Analysis*, 2nd edition, Springer-Verlag, 2008.
- [134] C. Sadosky, *Interpolation of Operators and Singular Integrals: An Introduction to Harmonic Analysis*, Dekker, 1979.
- [135] K. Saxe, *Beginning Functional Analysis*, Springer-Verlag, 2002.
- [136] R. Schilling, *Measures, Integrals, and Martingales*, Cambridge University Press, 2005.
- [137] A. Shiryaev, *Probability*, translated from the first (1980) Russian edition by R. Boas, 2nd edition, Springer-Verlag, 1996.
- [138] C. Silva, *Invitation to Ergodic Theory*, American Mathematical Society, 2008.

- [139] M. Simonnet, *Measures and Probabilities*, with a foreword by C.-M. Marle, Springer-Verlag, 1996.
- [140] Y. Sinai, *Introduction to Ergodic Theory*, translated by V. Scheffer, Princeton University Press, 1976.
- [141] Y. Sinai, *Probability Theory: An Introductory Course*, translated from the Russian and with a preface by D. Haughton, Springer-Verlag, 1992.
- [142] Y. Sinai, *Topics in Ergodic Theory*, Princeton University Press, 1994.
- [143] A. Skorokhod, *Basic Principles and Applications of Probability Theory*, edited by Y. Prokhorov, translated from the 1989 Russian original by B. Seckler, Springer-Verlag, 2005.
- [144] M. Steele, *The Cauchy–Schwarz Master Class*, Mathematical Association of America, Cambridge University Press, 2004.
- [145] E. Stein, *Topics in Harmonic Analysis Related to the Littlewood–Paley Theory*, Princeton University Press, 1970.
- [146] E. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, 1970.
- [147] E. Stein, *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals*, with the assistance of T. Murphy, Princeton University Press, 1993.
- [148] E. Stein and R. Shakarchi, *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*, Princeton University Press, 2005.
- [149] E. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton University Press, 1971.
- [150] R. Strichartz, *The Way of Analysis*, Jones and Bartlett, 1995.
- [151] K. Stromberg, *Introduction to Classical Real Analysis*, Wadsworth, 1981.
- [152] K. Stromberg, *Probability for Analysts*, lecture notes prepared by K. Ravindran, Chapman & Hall, 1994.
- [153] D. Stroock, *Probability Theory: An Analytic View*, Cambridge University Press, 1993.
- [154] D. Stroock, *A Concise Introduction to the Theory of Integration*, 3rd edition, Birkhäuser, 1999.
- [155] J. Taylor, *An Introduction to Measure and Probability*, Springer-Verlag, 1997.
- [156] M. Taylor, *Measure Theory and Integration*, American Mathematical Society, 2006.

- [157] A. Torchinsky, *Real Variables*, Addison-Wesley, 1988.
- [158] A. Torchinsky, *Real-Variable Methods in Harmonic Analysis*, Dover, 2004.
- [159] T. Tjur, *Probability Based on Radon Measures*, Wiley, 1980.
- [160] S. Varadhan, *Probability Theory*, American Mathematical Society, 2001.
- [161] S. Varadhan, *Stochastic Processes*, American Mathematical Society, 2007.
- [162] M. Väth, *Integration Theory: A Second Course*, World Scientific, 2002.
- [163] P. Walters, *An Introduction to Ergodic Theory*, Springer-Verlag, 1982.
- [164] R. Wheeden and A. Zygmund, *Measure and Integral: An Introduction to Real Analysis*, Dekker, 1977.
- [165] P. Whittle, *Probability via Expectation*, 4th edition, Springer-Verlag, 2000.
- [166] D. Williams, *Probability with Martingales*, Cambridge University Press, 1991.
- [167] P. Wojtaszczyk, *Banach Spaces for Analysts*, Cambridge University Press, 1991.
- [168] P. Wojtaszczyk, *A Mathematical Introduction to Wavelets*, Cambridge University Press, 1997.
- [169] K. Yosida, *Functional Analysis*, Springer-Verlag, 1995.
- [170] A. Zygmund, *Trigonometric Series*, Volumes I, II, 3rd edition, with a forward by R. Fefferman, Cambridge University Press, 2002.